

SID 1
best hit

only goes back
to

19 Aug
2000

RESULT 1
AC079031
LOCUS AC079031 187901 bp DNA linear HTG 02-OCT-2001
DEFINITION Homo sapiens chromosome 12q clone RP11-503G7, WORKING DRAFT
SEQUENCE, 7 unordered pieces.
ACCESSION AC079031
VERSION AC079031.15 GI:15809112
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 187901)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnini,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 187901)

1-1225

126-1557

AUTHORS Worley,K.C.
 TITLE Direct Submission
 JOURNAL Submitted (17-AUG-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 COMMENT On Sep 30, 2001 this sequence version replaced gi:14328945.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information
 Center project name: HBQO
 Center clone name: RP11-503G7

----- Summary Statistics
 Sequencing vector: Plasmid; M77789
 Sequencing vector: M13; L08821
 Chemistry: Dye-primer Bodipy: 32% of reads
 Chemistry: Dye-terminator Big Dye: 68% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 182139 bases at least Q40
 Consensus quality: 183938 bases at least Q30
 Consensus quality: 184894 bases at least Q20
 Estimated insert size: 185621; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 5.4x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 7 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 109303: contig of 109303 bp in length
 * 109304 109403: gap of unknown length
 * 109404 147116: contig of 37713 bp in length
 * 147117 147216: gap of unknown length
 * 147217 162615: contig of 15399 bp in length
 * 162616 162715: gap of unknown length
 * 162716 175476: contig of 12761 bp in length
 * 175477 175576: gap of unknown length
 * 175577 181264: contig of 5688 bp in length
 * 181265 181364: gap of unknown length
 * 181365 184738: contig of 3374 bp in length
 * 184739 184838: gap of unknown length
 * 184839 187901: contig of 3063 bp in length.

FEATURES Location/Qualifiers
 source 1. .187901
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="12q"
 /clone="RP11-503G7"

BASE COUNT 38363 a 55880 c 57328 g 35726 t 604 others

ORIGIN

Query Match 39.5%; Score 4067.2; DB 2; Length 187901;
 Best Local Similarity 87.5%; Pred. No. 0;
 Matches 4899; Conservative 0; Mismatches 61; Indels 637; Gaps 16;

```

Qy  3037  gggccttccccagggacagccgatgctctcctgatggctcctgcccttgacagagtgtctgc 3096
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 175585  GGGCCTCCCCCAGGTACAGCCGATGCTCTCCTGATGGCTCCTGCCCTTGCAGAGTGCTGC 175644

Qy  3097  ccccgctgcccacctggcctggaccctcgccctgagccccctcagggctctgcgccacct 3156
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 175645  CCCCGCTGCCACCTGGCCTGGACCCTCGCCTGAGCCCCCTCAGGGCTCTGCGCCACCT 175704

Qy  3157  caaccagggcgtttgttcgcaggaacctcccggtcttcccaactcgggaaaggaaggct 3216
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 175705  CAACCAGGCGTTTGTTCCGCAGGAACCTCCCGGCTCTTCCCACTCGGGAAAGGAAGGCT 175764

Qy  3217  ctgggcatggaggtcggccaggccccatccccgtaccctggcccttcttccctgcttcctg 3276
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 175765  CTGGGCATGGAGGTCGGCCAGGCCCATCCCCGTACCCTGCCCTTCTTCCCTGCTTCCTG 175824

Qy  3277  tttgtcactgccccggggcctttgcacctgcattccctctctctgtgagtgtcctggggc 3336
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 175825  TTTGTCACTGCCCCGGGGCCTTTGCACCTGCATTCCCTCTCTGTGAGTGTCTGGGGC 175884

Qy  3337  ccgttaccacgtcaccgtcccaggataccttttcttttcttctctctctccagcttta 3396
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 175885  CCGTTACCCACGTCACCGTCCCAGGATACCTTTTCTTTTCTTCTCTCTCTCCAGCTTTA 175944

Qy  3397  ttgaggtatagttgacaattcaggacgggtgtgcaactcaaggtatgcagcatcacacctg 3456
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 175945  TTGAGGTATAGTTGACAATTCAGGACGGTGTGCACTCAAGGTATGCAGCATCACACCTG 176004

Qy  3457  acacacgtaggcattgtgaaatgagtcccacaattgggctaattaacacaccccatcacct 3516
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 176005  ACACACGTAGGCATTGTGAAATGAGTCCCACAATTGGGCTAATTAACACACCCATCACCT 176064

Qy  3517  tacatggttacttcttctgtggtgagaacactaaattttaaatagaggacacacagcct 3576
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 176065  TACATGGTTACTTCTTCTGTGGTGAGAACTAAATTTTAAATAGAGGACACACAGCCT 176124

Qy  3577  gggcaacatagtgagaccctgtctctacaaatataaaaaattatctggacgtggtggtg 3636
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 176125  GGGCAACATAGTGAGACCCTGTCTCTACAAATATAAAAAAATTATCTGGACGTGGTGGTG 176184

Qy  3637  cacacctgtggtcccagctacttggaagctgaggctggagaatcacttgagcctggggag 3696
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 176185  CACACCTGTGGTCCCAGCTACTTGGGAAGCTGAGGCTGGAGAATCACTTGAGCCTGGGAG 176244

Qy  3697  gcggaggttgcggtgcaactccagcctggcgacagagggagggccctatctcaaaataaat 3756
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 176245  GCGGAGGTTGCGGTGCACTCCAGCCTGGGCGACAGAGGGAGG-CCTATCTCAAAATAAAT 176303

Qy  3757  aaataaaggacacattcttatcagctgtagtcaccacgttcattacatcttagaaccgc 3816
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

[illegible]

Qy 4106 ----- 4105

Db 177204 GTGTGTGTGTGTGTATGCATGTGTGTGTATTTCTTCCCTCCCCTCCCCCTCTCCTCC 177263

Qy 4106 -----ccacccttccctttccctcctgtggaataaacactcctgtgtgtgtgtgca 4156
 |||

Db 177264 CCTCCCTTCCCCACCCTTCCCTTTCCTCCTGTGGAATAAACTCCTGTGTGTGTGTGCA 177323

Qy 4157 tgcattgtgtgtgtatatttctcacatattttcattcatgcattcgttgatggaaacttgg 4216
 |||

Db 177324 TGCATGTGTGTGTATATTTCTCACATATTTTCATTCATGCATCCGTGATGGAACACTTGG 177383

Qy 4217 gttgattccgtgtcctggctgctgggacagtgtgctgcgatgaacacgaggggtacagacgcc 4276
 |||

Db 177384 GTTGATTCCGTGTCTTGGCTGCTGGGACAGTGTGCGATGAACACGAGGGTACAGACGCC 177443

Qy 4277 tctcctacacgctaatttcaactccttggatatacaccagcagtgaggattgctggatca 4336
 |||

Db 177444 TCTCCTACACGCTAATTTCAACTCTTGGATATACACCAGCAGTGGGATTGCTGGATCA 177503

Qy 4337 ggtgggagctctatttccacatttttgaggaaacctccctgccgtctcccatggtggctgt 4396
 |||

Db 177504 GGTGGGAGCTCTATTTCCACATTTTGGAGAACCTCCCTGCCGTCTCCCATGGTGGCTGT 177563

Qy 4397 gccaacgacgttcccagggacagagtgaacgggccccttccctccatgtcctcgccaac 4456
 |||

Db 177564 GCCAACGACGTTCCCAGGGACAGAGTGCAACGGGCCCTTTCCTCCATGTCTCGCCAAC 177623

Qy 4457 actcgctatcttttgcgttttgatgacagtcacccaataggtgccagttggtacctcct 4516
 |||

Db 177624 ACTCGCTATCTTTTGCCTTTTGATGACAGTCATCCCAATAGGTGCCAGTTGGTACCTCCT 177683

Qy 4517 gtgggtttttatttgattttcctgatgattagtgatgctggacgttatttcgtctacactt 4576
 |||

Db 177684 GTGGTTTTTATTTGATTTTCTGATGATTAGTGATGCTGGACGTTATTTCTGTCTACACTT 177743

Qy 4577 cggccacttacatgttttccttcgagacacgcagattcaggtcctttgcacgttttaaaa 4636
 |||

Db 177744 CGGCCACTTACATGTTTTCTTCGAGACACGCAGATTAGGTCTTTGCACGTTTTAAAA 177803

Qy 4637 ttttttttgtttgtttttgattgagttgaattccttctacaatttgcaaattaactcc 4696
 |||

Db 177804 TTTTTTTGTTTGTGTTTTGTTATTGAGTTGAATTCCTTCTACAATTTGCAAATTAACCTCC 177863

Qy 4697 tcatcatatacatggattgcaaatacccccgctccccctgggggttttgcttttctactg 4756
 |||

Db 177864 TCATCATATACATGGATTGCAAATACCCCCGCTCCCCCTGGGGTTTTGCCTTTTCACTG 177923

Qy 4757 caaatactcccgctccccatgggggttgcttttccctgccaatacccccaactcccca 4816
 |||

Db 177924 CAAATACTCCCGCTCCCCATGGGGGTGCTTTTCCCTGCCAATACCCCCACCTCCCCA 177983

Qy 4817 tgggggttgcttttccctgcaaatacccccaactccccgtgggttctgccttttccctg 4876
 |||

Db 177984 TGGGGGTGCTTTTCCCTGCAAATACCCCCACCTCCCCGTGGGTCTGCCTTTTCCCTG 178043

Qy 4877 ccaatacccccgcctccccctgggggtgccttttcaactctgttggtttcctttgcggaa 4936
 |||
 Db 178044 CCAATACCCCCGCCTCCCCCTGGGGGTGCCTTTTCACTCTGTTGGTTTCCTTTCGGGAA 178103

Qy 4937 gctttctggtttgttgactctcactgtctatttttgccttctgttgctgtgcttgtggg 4996
 |||
 Db 178104 GCTTTCTGGTTTGTGCACTCTCACTGTCTATTTTGCTTCTGTTGCCTGTGCTTGTGGG 178163

Qy 4997 gccatattttaaaaaaatcattgcccgaccagcctcaagaagttttcctcctaagtttt 5056
 |||
 Db 178164 GCCATATTTTAAAAAATCATTGCCCGGACCAGCCTCAAGAAGTTTTCCTCCTACGTTTT 178223

Qy 5057 cttctaagagttttatggtgtcgggtcttaggtttgaatctttaatccgtgttgagttga 5116
 |||
 Db 178224 CTTCTAAGAGTTTTATGGTGTGCGGTCTTAGGTTTGAATCTTTAATCCGTGTTGAGTTGA 178283

Qy 5117 ttttcgtaggtggtgtcggatgaggccctttcatcctcctccacttttccagcaccacc 5176
 |||
 Db 178284 TTTTCGTAGGTGGTGTGCGATGAGGCCCTTTCATCCTCCTCCACTTTTCCAGCACCACC 178343

Qy 5177 tattgaggatgccccctttccccgtcgtgtgtccttggcgcctttgctgaaggtcagttgg 5236
 |||
 Db 178344 TATTGAGGATGCCCTTTCCCCGTCTGTGTCTTGGCGCCTTTGCTGAAGGTCAGTTGG 178403

Qy 5237 ccgtaactgtgcatggggacccttcctggccccctggtgccctgtgccccatagtccc 5296
 |||
 Db 178404 CCGTAACTGTGCATGGGGACCCTTCCTGGCCCCCTGGTGCCCTGTGCCCCATATGTCCC 178463

Qy 5297 accccctcccttactttttctccatggcatgaatcacccagacctatacaaaattt 5356
 |||
 Db 178464 ACCCCCTCCCTTACTTTTTCTCCATGGCATGAATCACCCAGACCTACTATACATATTTT 178523

Qy 5357 atcctattttatttttattttattttatttttggagatggagtctcactctgtcaccag 5416
 |||
 Db 178524 ATCTTATTTATTTTATTTATTTATTTATTTTGGAGATGGAGTCTCACTCTGTCACCAG 178583

Qy 5417 gctggagtgcagtgggcacgatctcggtcactgcaagctccgcctcccaggttcacgcca 5476
 |||
 Db 178584 GCTGGAGTGCAGTGGCAGATCTCGGCTCACTGCAAGCTCCGCCTCCCAGGTTACGCCA 178643

Qy 5477 ttctcctgcctcagcctcccaagtagctgggattacgggcgcccgccaccatgccggct 5536
 |||
 Db 178644 TTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGATTACGGGCGCCCGCCACCATGCCGGCT 178703

Qy 5537 aaattttttgtttttttcgtagagacagagtttccctatgttgccccaggttggtctcg 5596
 ||
 Db 178704 AATTTTTTTGTATTTTGTAGTGAGACGGAGTTTCAACATGTT--AGCCAGGATGGTCTCG 178761

Qy 5597 aactcctgggctcaagtaatccttccacttcggcctcccaaagtgcgtgggattacaggca 5656
 ||
 Db 178762 ATCTCCTGACCTC--GTGATCCACCCGCCTCAGCCTCCCAAAGTGCTGGGATTATAGGCG 178819

Qy 5657 tgagccattcggcccgccctattttttttttttcagacagagtttactctgtcaccaca 5716
 |||
 Db 178820 TGAGCCATCGCGCCCGCCTATTTTTTTTTTTTTCAGACAGAGTTTCACTCTTGTACCCCA 178879

Qy 5717 ggctgaaatgcattgcaatgatcttggtcactgcaacttccacctcccaggttcaaagg 5776

Db 178880 GGCTGGAGTGCAGTGGCATGATCTTGGCTCACTGCAACTTCCACCTCCCAGGTTCAAGCG 178939
 Qy 5777 atttttctggcctcagcctccccgaggagctgggattacagtgtgcaaccaccacaccggg 5836
 Db 178940 ATTCTCCT-GCCTCAGCCTCCCGAGTAGCTGGGATTACAGTGTGC-ACCACCACACCTGG 178997
 Qy 5837 ctaaaatttttgaatttttttttttactagagacagggttcaacaatgctggtcaggc 5896
 Db 178998 CT-AAATTTTGTATTTTTTTTTTTTACTAGAGACAGGGTTTCAACATGCTGGTCAGGC 179056
 Qy 5897 tggctctgaattcctgacctaagtgatecctccacctcggcctcccaaagtgtgggat 5956
 Db 179057 TGGTCTCGAATTCTGACCTCAAGTGATCCTCCACCTCGGCCTCCCAAAGTGCTGGGAT 179116
 Qy 5957 tacaggcgtgagccgcatgcctggccatggatattgtaaagtgtcttgtttgtgtatg 6016
 Db 179117 TACAGGCGTGAGCCGCCATGCCTGGCCATGGATATTGTAAATGTTCTTGTGTGTATG 179176
 Qy 6017 ttttcctcactgggctgtgcactcctgagggcggggcatctgtcccattcttcagtgtg 6076
 Db 179177 TTTTCCTCACTGGGCTGTGCACTCCTGAGGGCGGGGCATCTGTCCCATTCTTCAGTGCTG 179236
 Qy 6077 ggtccctgtgtctgggacagtgtatacatagcaggtgcataatcagtcttgactgga 6136
 Db 179237 GGTCCCTGTGTCTGGGACAGTGTATACATACAGCAGGTGCATAATCAGTCTTGACTGGA 179296
 Qy 6137 aggggtgaggagtgcaacgcacatggcagtcattggactatgtgtctgagaagcataactc 6196
 Db 179297 AGGGTGAGGGAGTCAACGCACATGGCAGTCATTGGACTATGTGTCTGAGAAGCATAACTC 179356
 Qy 6197 acttaatcttgaagttcacttatggattgaagtgtgcggttcagtgacttttaatatatt 6256
 Db 179357 ACTTAATCTTGAAGTTCACCTATGGATTGAAGTGTGCGGTTCAAGTACTTTTAATATATT 179416
 Qy 6257 taccgagttgtgtaaccatcaccaccatctaatttttaaatcattttcatcatccctaaaa 6316
 Db 179417 TACCGAGTTGTGTAACCATCACACCATCTAATTTTAAATCATTTTCATCATCCCTAAAA 179476
 Qy 6317 gaaacttcagaccactagctgtccctccccctattcctccaccccagccctggtcctg 6376
 Db 179477 GAAACTTCAGACCCACTAGCTGTCCCTCCCCCTATTCTCCACCCCAGCCCTGGTCCTG 179536
 Qy 6377 gccgcaggctgtcacctgcatctctctgtggatctgccggttggtggacatttcacacac 6436
 Db 179537 GCCGCAGGCTGCTCACCTGCATCTCTCTGTGGATCTGCCGGTGTGGACATTTCACACAC 179596
 Qy 6437 ctgctgtcagctcttctgtgcctgcctctttcactcgtgtgatgtttaagttcaccatg 6496
 Db 179597 CTGCGTGCAGTCTTCTGTGCCTGCCCTTTCACTCGCTGTGATGTTTAAGTCACCCATG 179656
 Qy 6497 ttgtcatctatatcggtacttacttccttttttttttggagatgaagtcttgctcttgt 6556
 Db 179657 TTGTCATCTATATCGGGACTTACTTCCTTTTTTTTTTTGGAGATGAAGTCTTGCTCTGT 179716
 Qy 6557 caccaggtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtg 6616

Db 179717 CACCCAGGCTGGAGTGCAGTGGCGTGATCTCGGCTCACAGCAACTTCTGCCTCTGGGGTT 179776

Qy 6617 caagtgattctcctgccttagcctcccaagtagctgggactacaggtttgcaccaccatg 6676
 |||

Db 179777 CAAGTGATTCTCCTGCCTTAGCCTCCCAAGTAGCTGGGACTACAGGTTTGCACCACCATG 179836

Qy 6677 tcctgctaattttttttttttttgtatttttaatagagacagggtttctcctcattggcca 6736
 |||

Db 179837 TCCTGCTAATTTTTTTTTTTTTTGTATTTTAAATAGAGACAGGGTTTCTCCTCATTGGCCA 179896

Qy 6737 ggctgggtctcgaactcctgacctcagacgatccacctgcctcagcctcccgaagtgttg 6796
 |||

Db 179897 GGCTGGTCTCGAACTCCTGACCTCAGACGATCCACCTGCCTCAGCTCCCGAASTGTTGG 179956

Qy 6797 gattacaggcacgagccactgtgcccggccatcattccttttactgtgactaatagtc 6856
 |||

Db 179957 GATTACAGGCACGAGCCACTGTGCCCGGCCATCATTCTTTTACTGCTGACTAATAGTC 180016

Qy 6857 tgctgtgtgaatccaccgctagaaacccactcatcagttgatggcatgtgggttgcctc 6916
 |||

Db 180017 TGCTGTGTGAATCCACCGCTAGAAACCCACTCATCAGTTGATGGTCATGTGGGTGCTTC 180076

Qy 6917 tgctattcgcttattatgaacagtgtggaataaacgttcctgtgcactcttgggcatac 6976
 |||

Db 180077 TGCTATTCGCTTATTATGAACAGTGCTGGAATAAACGTTCTGTGCACTCTTGGGCATAC 180136

Qy 6977 gcctaggagtggaaactgctgggtcaaagtgtgactttacgtttaacgttctgaggagccg 7036
 |||

Db 180137 GCCTAGGAGTGGAAGTCTGGGTCAAATGGTGACTTTACGTTTAACTGCTGAGGAGCCG 180196

Qy 7037 ccaggcgttttaacacagtgcactgcaccatttcacattcctgccaacaatgtgtgagaat 7096
 |||

Db 180197 CCAGGCGTTTAAACACAGTGACTGCACCATTTCACATTCTGCCAACAATGTGTGAGAAT 180256

Qy 7097 tccaatttctctacatccccaacattttccttttaaaaaaagaaaaaagaaacatagcca 7156
 |||

Db 180257 TCCAATTTCTCTACATCCCAACATTTCTTTTAAAAAAGAAAAAAGAAACATAGCCA 180316

Qy 7157 tctaagtggatgtggagcagactgtccctctgggtttgggtttgcgttgcttttatggctc 7216
 |||

Db 180317 TCTAAGTGGATGTGGAGCAGACTGTCCCTCTGGTTGGGTTGCGTTGCTTTTATGGCTC 180376

Qy 7217 atgatgtctgagtctctctccatgtgctcatggggattcgatatctactttgggaaatg 7276
 |||

Db 180377 ATGATGTCTGAGTCTCTCTCCATGTGCTCATGGGATTCTGTATATCTACTTTGGGAAATG 180436

Qy 7277 cttattcaagtcctttgtccacatttgactgggttgcttgcctttttatttcatttacta 7336
 |||

Db 180437 CTTATTCAAGTCCTTTGTCCACATTGACTGGGTTGCTTGTCTTTTATTTTACTTACTA 180496

Qy 7337 cgatgacagcccctacatggaaggattttgtttttgtaatcccattaccccgagggtgaga 7396
 |||

Db 180497 CGATGACAGCCCCTACATGGAAGGATTTTGTTTTGTAAATCCCATTACCCCGAGGTGAGA 180556

Qy 7397 atgaattgccagttgctcaaggccttcagctcttagggaggagcctggacctggagctgc 7456
 |||

Db 180557 ATGAATTGCCAGTTGCTCAAGGCCTTCAGCTCTTAGGGAGGAGCCTGGACCTGGAGCTGC 180616

Qy 7457 tccgggctctggcaaagctccaatcccgccctcagtccttgaggcctggctcctcaccag 7516
 |||||||
 Db 180617 TCCGGGCTCTGGCAAAGCTCCAATCCCGGCCCTCAGTCCTTGAGGCCTGGTCCTCACCCAG 180676

 Qy 7517 ctttctccttcaccgtgccatggaggaagcccgacctccctgcacggctggcctgggg 7576
 |||||||
 Db 180677 CTTTCTCCTTCCACCGTGCCATGGAGGAAGCACGACCTCCCTGCACGGCTGGCCTGGGGT 180736

 Qy 7577 tgttcacgactgagtccaggtgtccccagaacggatgtcactggtcacagtgttcctgg 7636
 |||||||
 Db 180737 TGTTACGACTGAGTCCAGGTGTCCCCAGAACGGATGTCACTGGTCACAGTGTTCCTGGT 180796

 Qy 7637 aataggtgacccccaggcacaggggtgttcctgatcataggttaaccaggcacaggtgtccc 7696
 |||||||
 Db 180797 AATAGGTGA-CCCAGGCACAGGGTGTTCCTGATCATAGGTAACCCAGGCACAGGTGT-CC 180854

 Qy 7697 agtcacaggtgtctccaggcacaggtgtccccagtcacaggtgtcccaggtcacaggcgt 7756
 |||||||
 Db 180855 AGTCACAGGTGTCTCCAGGCACAGGTGT-CCCAGTCACAGGTGT-CCAGGTCACAGGCGT 180912

 Qy 7757 ccccaggcacaggtgtccctggtcacagatgtccccaggcacaggtgtcccaggcacagg 7816
 |||||||
 Db 180913 -CCCAGGCACAGGTGT-CCTGGTCACAGATGT-CCCAGGCACAGGTGT-CCAGGCACAGG 180968

 Qy 7817 tgtctccaggcacaggcgtcccaggtcacaggtgtccccggtcacaggtgtccctgg 7876
 |||||||
 Db 180969 TGTCTCCAGGCACAGGCGTCCCAGGTACAGGTGTNCCCGGTCACAGGTGTCCCTGGTCA 181028

 Qy 7877 caggtgtctccaggcacaggtgtccctggtcacaggtgtccccggtcacaggtgtcccag 7936
 |||||||
 Db 181029 CAGGTGTCTCCAGGCACAGGTGTNCCTGGTCACAGGGTGTCCCGGTCACAGGTGTCCCAG 181088

 Qy 7937 gtcacaggtgtccccaggcacaggtgtccccggtcacaggtgtctccggtcacaggtgtc 7996
 |||||||
 Db 181089 GTCACAGGTGTNCCAGGCACAGGTGTCCCCGGTCACAGGTGTCTCCGGTCACAGGTGT- 181147

 Qy 7997 cccaggcataggtgtcc 8013
 |||||||
 Db 181148 CCCAGGCATAGGTGTCC 181164

all of SID 64

RESULT 1
AC079031/c
LOCUS AC079031 187950 bp DNA HTG 26-JUN-2001
DEFINITION Homo sapiens chromosome 12q clone RP11-503G7, WORKING DRAFT
SEQUENCE, 8 unordered pieces.
ACCESSION AC079031
VERSION AC079031.14 GI:14328945
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 187950)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 187950)
AUTHORS Worley,K.C.
TITLE Direct Submission

JOURNAL Submitted (17-AUG-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Jun 8, 2001 this sequence version replaced gi:13794207.

----- Genome Center

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: HBQO
Center clone name: RP11-503G7

----- Summary Statistics

Sequencing vector: Plasmid; M77789
Sequencing vector: M13; L08821
Chemistry: Dye-primer Bodipy: 36% of reads
Chemistry: Dye-terminator Big Dye: 64% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 179703 bases at least Q40
Consensus quality: 182695 bases at least Q30
Consensus quality: 184221 bases at least Q20
Estimated insert size: 184740; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 5.1x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 56563: contig of 56563 bp in length
* 56564 56663: gap of unknown length
* 56664 92323: contig of 35660 bp in length
* 92324 92423: gap of unknown length
* 92424 121410: contig of 28987 bp in length
* 121411 121510: gap of unknown length
* 121511 150219: contig of 28709 bp in length
* 150220 150319: gap of unknown length
* 150320 168275: contig of 17956 bp in length
* 168276 168375: gap of unknown length
* 168376 181133: contig of 12758 bp in length
* 181134 181233: gap of unknown length
* 181234 184787: contig of 3554 bp in length
* 184788 184887: gap of unknown length
* 184888 187950: contig of 3063 bp in length.

FEATURES Location/Qualifiers
source 1. .187950
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12q"
/clone="RP11-503G7"

BASE COUNT 37588 a 56082 c 57191 g 36388 t 701 others

ORIGIN

Query Match 100.0%; Score 40; DB 2; Length 187950;
Best Local Similarity 100.0%; Pred. No. 1.7e-05;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 actacagggttgaccaccatgtcctgctaattttttttt 40
|||||
Db 160697 ACTACAGGTTTGCACCACCATGTCCTGCTAATTTTTTTT 160658

SID1 good for oligos.

RESULT 3
HSU14383
LOCUS HSU14383 1403 bp mRNA linear PRI 31-DEC-1994
DEFINITION Human mucin (MUC8) mRNA, partial cds.
ACCESSION U14383 U04799
VERSION U14383.1 GI:606953
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 907 to 1403)
AUTHORS Sachdev,G.P.
TITLE Carboxy terminus of a major human tracheobronchial mucin MUC8
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 941)
AUTHORS Shankar,V., Gilmore,M.S., Elkins,R.C. and Sachdev,G.P.
TITLE A novel human airway mucin cDNA encodes a protein with unique
tandem-repeat organization
JOURNAL Biochem. J. 300 (Pt 2), 295-298 (1994)
MEDLINE 94271137
REFERENCE 3 (bases 1 to 1403)
AUTHORS Sachdev,G.P.
TITLE Direct Submission
JOURNAL Submitted (03-SEP-1994) Goverdhan P. Sachdev, University of
Oklahoma Health Sciences Center, Medicinal Chemistry, College of
Pharmacy, 1110 N. Stonewall Avenue, Oklahoma City, OK 73117, USA
COMMENT On Jan 1, 1995 this sequence version replaced gi:537430.
FEATURES
source Location/Qualifiers
1. .1403
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="PAM2"
/tissue_type="trachea/bronchus"
/dev_stage="adult"
repeat_region <1. .892
/rpt_type=tandem
/rpt_unit=3. .43
gene 1. .1403
/gene="MUC8"
CDS <1. .944
/gene="MUC8"
/codon_start=3
/product="mucin"
/protein_id="AAA58346.1"
/db_xref="GI:501033"
/translation="TSCPRPLQEGTPGSRAAHALSRRGHRVHELPTSSPGGDTGFMSC
PRPFQEGTPGSRAAHVLSRKGPRVHELPTSSPGRDPGFTSCPRPLQEGTRVTNCPRL
QEGTPGSRAAHVLSRRGHRVHELPTSPGRDPGFMSCPRPLQEGTRVTNCPRLQEGT
RVTSCPRRLQEGTRVTSCPRPLQEGTRVTNCPRALQEGTPGSRAAHALSRRGPRVHEL
PTSSPGGDTGFTSCPRPLQEGTPGSRAAHALSRRGHRVHELPTSSPGRDPGHELPTSS
PGGDTGFTSCPRTFQEGTPGSGLLPAAHIVPLCKSEER"
3'UTR 945. .1403
/gene="MUC8"
polyA_signal 1350. .1355
/gene="MUC8"

polyA_site 1403
/gene="MUC8"
/note="17 A residues"
BASE COUNT 254 a 490 c 413 g 246 t
ORIGIN

Query Match 9.4%; Score 964.6; DB 9; Length 1403;
Best Local Similarity 95.6%; Pred. No. 2.1e-137;
Matches 1100; Conservative 0; Mismatches 39; Indels 12; Gaps 10;

```
Qy 1482 cacgagctgcccacgtcctctccaggaagggaccccggttcacgagctgcccacgtcgt 1541
      |||
Db 2 CACGAGCTGCCCACGTCCTCTCCAGGAGGGGACACCGGGTTCACGAGCTGCCCACGCCCT 61

Qy 1542 ctccaggaagggac-cgggtccacgagctgcccacgtcctctccaggaagggac-cgg 1599
      |||
Db 62 CTCCAGGAGGGGACACCGGGTTCACGAGCTGCCCACGTCCTCTCCAGGAGGGGACACCGG 121

Qy 1600 gtccacgagctggccacgtcctctgcaggaagggaccccggttcacgagctgcccacgt 1659
      ||
Db 122 GTTCATGAGCTGCCCACGCCCTTTCCAGGAAGGGACCCCGGGTTCACGAGCTGCCCACGT 181

Qy 1660 cctctccaggaagggaccccggttcacgagctgcccacgtcctctccaggaagggaccc 1719
      |||
Db 182 CCTCTCCAGGAAGGGACCCCGGGTTCACGAGCTGCCCACGTCCTCTCCAGGAAGGGACCC 241

Qy 1720 cgggtccacgagctgcccacgtcctctccaggaagggaccccggttcacgaactgccc 1779
      |||
Db 242 CGGGTTCACGAGCTGCCCACGTCCTCTCCAGGAAGGGACCC--GGGTCACGAAGTGCCCA 299

Qy 1780 cgtcctctccaggaagggaccccggttcacgagctgcccacgtcctctccaggaaggg 1839
      |||
Db 300 CGTCCTCTCCAGGAAGGGACCCCGGGTTCACGAGCTGCCCACGTCCTCTCCAGGAGGGGA 359

Qy 1840 caccgggttcacgagctgcccacgcctctccaggaagggaccccggttcacgagctgc 1899
      |||
Db 360 CACCGGGTTCACGAGCTGCCCACGCCCTCTCCAGGAAGGGACCCCGGGTTCATGAGCTGC 419

Qy 1900 ccacgtcctctccaggaagggaccccggttcacgaactgcccacgcctctccaggaaggg 1959
      |||
Db 420 CCACGTCCTCTCCAGGAAGGGACCCGGGT-CACGAAGTCCCACGCCCTCTCCAGGAGGG 478

Qy 1960 gacccggttcacgagctgcccacgtcgtcaacgggaagggaccccggttcacgagctgc 2019
      |||
Db 479 GACCCGGGT-CACGAGCTGCCCACGTCGTCTCCAGGAAGGGACCCGGGT-CACGAGCTGC 536

Qy 2020 ccacgtcctctccaggaagggaccccggttcacgaactgcccacgcgtctccaggaaggg 2079
      |||
Db 537 CCACGTCCTCTCCAGGAAGGGACCCGGGT-CACGAAGTCCCACGCGCTCTCCAGGAGGG 595

Qy 2080 gacacgggttcacgagctgcccacgcctctccaggaagggaccccggttcacgagct 2139
      |||
Db 596 GACACCGGGTTCACGAGCTGCCCACGCCCTCTCCAGGAAGGGACCCCGGGTTCACGAGCT 655

Qy 2140 gcccacgtcctctccaggaagggaccccggttcacgagctgcccacgtcctctccagga 2199
      |||
```

Db 656 GCCCACGTCCTCTCCAGGAGGGGACACCGGGTTCACGAGCTGCCACGTCCTCTCCAGGA 715
 Qy 2200 ggggacaccgggttcacgagctgccacgcccctctccaggaggggacaccgggttcacga 2259
 |||||
 Db 716 GGGGACACCGGGTTCACGAGCTGCCACGCCCTCTCCAGGAGGGGACACCGGGTTCACGA 775
 Qy 2260 gctgcccacgtcctctccaggaagggacccgggtccacgagctgccacgtcctctccag 2319
 |||||
 Db 776 GCTGCCACGTCCTCTCCAGGAAGGGACCCGGGT-CACGAGCTGCCACGTCCTCTCCAG 834
 Qy 2320 gaggggacaccgggttcacgagctgccacgcactttccaggaagggaccccggttcag 2379
 |||||
 Db 835 GAGGGGACACCGGGTTCACGAGCTGCCACGCACCTTCCAGGAAGGGACCCCGGGTTCAG 894
 Qy 2380 gtctcctgcccggcccacatcgctgcctttgtgtaaatcagaagaaagatgaggaacaggcc 2439
 |||||
 Db 895 GTCTCCTGCCGGCCACATCGTGCCTTTGTGTAAATCAGAAGAAAGATGAGGAACAGGCC 954
 Qy 2440 ctctctctctccaggcaggctttggtggaggggctggatctcctgccgcaccttccctg 2499
 |||||
 Db 955 CTCCTCTCTCTCCAGGCAGGCTTTGGTGGAGGGGCTGGATCTCCTGCCGCACCTTCCCTG 1014
 Qy 2500 gcagggcaccctgtgcttgagccccagaactgcaggcgccggcagagaaggggtccatg 2559
 ||| |||||
 Db 1015 GCA-GGCACCCTGTCGTTGAGCCCCAGAACTGCAGGCGGCCGGCAGAGAAGGGGTCCATG 1073
 Qy 2560 atggcgcctcggtgcgacgcttgacctgccccatggacctgggtgaggacttcccag 2619
 ||||| |||||
 Db 1074 ATGGCGCCTCGGTGCG--GCCTTGGACCTGCCCCATGGACCTGGAGACAGGGTTTCTCC 1131
 Qy 2620 cccttccccgg 2630
 | || || |
 Db 1132 TCATTGGCCAG 1142

Full sid 2

HSU14383
LOCUS HSU14383 1403 bp mRNA linear PRI 31-DEC-1994
DEFINITION Human mucin (MUC8) mRNA, partial cds.
ACCESSION U14383 U04799
VERSION U14383.1 GI:606953
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 907 to 1403)
AUTHORS Sachdev,G.P.
TITLE Carboxy terminus of a major human tracheobronchial mucin MUC8
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 941)
AUTHORS Shankar,V., Gilmore,M.S., Elkins,R.C. and Sachdev,G.P.
TITLE A novel human airway mucin cDNA encodes a protein with unique
tandem-repeat organization
JOURNAL Biochem. J. 300 (Pt 2), 295-298 (1994)
MEDLINE 94271137
REFERENCE 3 (bases 1 to 1403)
AUTHORS Sachdev,G.P.
TITLE Direct Submission
JOURNAL Submitted (03-SEP-1994) Goverdhan P. Sachdev, University of
Oklahoma Health Sciences Center, Medicinal Chemistry, College of
Pharmacy, 1110 N. Stonewall Avenue, Oklahoma City, OK 73117, USA
COMMENT On Jan 1, 1995 this sequence version replaced gi:537430.
FEATURES
source Location/Qualifiers
1. .1403
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="PAM2"
/tissue_type="trachea/bronchus"
/dev_stage="adult"
repeat_region <1. .892
/rpt_type=tandem
/rpt_unit=3. .43
gene 1. .1403
/gene="MUC8"
CDS <1. .944
/gene="MUC8"
/codon_start=3
/product="mucin"
/protein_id="AAA58346.1"
/db_xref="GI:501033"
/translation="TSCPRPLQEGTPGSRAAHALSRRGHRVHELPTSSPGGDTGFMS
CPRPFQEGTPGSRAAHVLSRKGPVHELPTSSPGRDPGFTSCFRPLQEGTRVTNCPRPL
QEGTPGSRAAHVLSRRGHRVHELPTSPGRDPGFMSCPRLQEGTRVTNCPRPLQEGT
RVTSCPRRLQEGTRVTSCPRPLQEGTRVTNCPRALQEGTPGSRAAHALSRRGPRVHEL
PTSSPGGDTGFTSCPRPLQEGTPGSRAAHALSRRGHRVHELFTSSPGRDPGHELPTSS
PGGDTGFTSCPRTFQEGTPGSGLLPAHIVPLCKSEER"
3'UTR 945. .1403
/gene="MUC8"
polyA_signal 1350. .1355
/gene="MUC8"
polyA_site 1403

/gene="MUC8"
/note="17 A residues"

BASE COUNT 254 a 490 c 413 g 246 t
ORIGIN

Query Match 67.8%; Score 1072.4; DB 9; Length 1403;
Best Local Similarity 88.4%; Pred. No. 1.9e-225;
Matches 1373; Conservative 0; Mismatches 26; Indels 155; Gaps 13;

```
Qy      2  cacgagctgcccacgtcctctccaggaagggaccccggttcacgagctgcccacgtcgt 61
      |||
Db      2  CACGAGCTGCCCACGTCCTCTCCAGGAGGGGACACCGGGTTCACGAGCTGCCCACGCCCT 61

Qy     62  ctccaggaagggac-ccgggtccacgagctgcccacgtcctctccaggaagggac-ccgg 119
      |||
Db     62  CTCCAGGAGGGGACACCGGGTTCACGAGCTGCCCACGTCCTCTCCAGGAGGGGACACCGG 121

Qy    120  gtccacgagctggccacgtcctctgcaggaagggaccccggttcacgagctgcccacgt 179
      ||
Db    122  GTTCATGAGCTGCCCACGCCCTTTCAGGAAGGGACCCCGGGTTCACGAGCTGCCCACGT 181

Qy    180  cctctccaggaagggaccccggttcacgagctgcccacgtcctctccaggaagggaccc 239
      |||
Db    182  CCTCTCCAGGAAGGGACCCCGGGTTCACGAGCTGCCCACGTCCTCTCCAGGAAGGGACCC 241

Qy    240  cgggtccacgagctgcccacgtcctctccaggaagggaccccggttcacgaactgcca 299
      |||
Db    242  CGGGTTCACGAGCTGCCCACGTCCTCTCCAGGAAGGGACCC--GGGTCACGAAGTGCCCA 299

Qy    300  cgtcctctccaggaagggaccccggttcacgagctgcccacgtcctctccaggagggga 359
      |||
Db    300  CGTCCTCTCCAGGAAGGGACCCCGGGTTCACGAGCTGCCCACGTCCTCTCCAGGAGGGGA 359

Qy    360  caccgggttcacgagctgcccacgcctctccaggaagggaccccggttcacgagctgc 419
      |||
Db    360  CACCGGGTTCACGAGCTGCCCACGCCCTCTCCAGGAAGGGACCCCGGGTTCATGAGCTGC 419

Qy    420  ccacgtcctctccaggaagggaccccggttcacgaactgcccacgcctctccaggaggg 479
      |||
Db    420  CCACGTCCTCTCCAGGAAGGGACCCGGGT-CACGAAGTGCCACGCCCTCTCCAGGAGGG 478

Qy    480  gacccggttcacgagctgcccacgtcgtcaacgggaagggaccccggttcacgagctgc 539
      |||
Db    479  GACCCGGGT-CACGAGCTGCCCACGTCTCCAGGAAGGGACCCGGGT-CACGAGCTGC 536

Qy    540  ccacgtcctctccaggaagggaccccggttcacgaactgcccacgcgctctccaggaggg 599
      |||
Db    537  CCACGTCCTCTCCAGGAAGGGACCCGGGT-CACGAAGTGCCACGCGCTCTCCAGGAGGG 595

Qy    600  gacacgggttcacgagctgcccacgcctctccaggaagggaccccggttcacgagct 659
      |||
Db    596  GACACCGGGTTCACGAGCTGCCCACGCCCTCTCCAGGAAGGGACCCCGGGTTCACGAGCT 655

Qy    660  gcccacgtcctctccaggagggacaccgggttcacgagctgcccacgtcctctccagga 719
      |||
Db    656  GCCCACGTCCTCTCCAGGAGGGGACACCGGGTTCACGAGCTGCCCACGTCCTCTCCAGGA 715
```

Qy 720 ggggacaccgggttcacgagctgccacgcccctctccaggaggggacaccgggttcacga 779
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 716 GGGGACACCGGGTTCACGAGCTGCCACGCCCTCTCCAGGAGGGGACACCGGGTTCACGA 775

 Qy 780 gctgccacgctcctctccaggaagggacccgggtccacgagctgccacgctcctctccag 839
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 776 GCTGCCACGTCTCTCCAGGAAGGGACCCGGGT-CACGAGCTGCCACGTCTCTCCAG 834

 Qy 840 gaggggacaccgggttcacgagctgccacgcactttccaggaagggaccccggttcag 899
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 835 GAGGGGACACCGGGTTCACGAGCTGCCACGCACTTTCAGGAAGGGACCCCG3GTTTCAG 894

 Qy 900 gtctcctgcccggcccacatcgctgcctttgtgtaaatacagaagaaagatgaggaacaggcc 959
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 895 GTCTCCTGCCGGCCACATCGTGCCTTTGTGTAAATCAGAAGAAAGATGAGGAACAGGCC 954

 Qy 960 ctctctctctctccaggcaggctttgggtggaggggctggatctcctgccgcaccttcctg 1019
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 955 CTCCTCTCTCTCCAGGCAGGCTTTGGTGGAGGGGCTGGATCTCCTGCCGCACCTTCCTG 1014

 Qy 1020 gcagggcaccctgtgcttgagccccagaactgcaggcgccggcagagaaggggtccatg 1079
 ||| ||||||||||| ||||||||||||||||||||||||||||||||||||||||
 Db 1015 GCA-GGCACCCTGTCTGTTGAGCCCCAGAACTGCAGGCGGCCGGCAGAGAAGGGGTCCATG 1073

 Qy 1080 atggcgccctcggtgcgcagccttggacctgccccatggacctgggaacctcccggctct 1139
 ||||||||||||||| ||||||||||||||||||||||||||||||||||||
 Db 1074 ATGGCGCCTCGGTGCG--GCCTTGGACCTGCCCCATGGACCTGG----- 1116

 Qy 1140 tcccactcgggaaaggaaggctctgggcatggaggtcggccaggcccatcccgtaccc 1199
 Db 1117 ----- 1116

 Qy 1200 tggcccttcttctcctgcttcctggttgcactgccccggggcctttgcacctgcattcct 1259
 Db 1117 ----- 1116

 Qy 1260 ctctctagacagggtttctcctcattggccaggctgggtctcgaactcctgacctcagacg 1319
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1117 -----AGACAGGGTTCTCCTCATTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGACG 1170

 Qy 1320 atccacctgcctcagcctcccgaagtgttgggattacaggcacgagccactgtgcccggc 1379
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1171 ATCCACCTGCCTCAGCCTCCCGAAGTGTTGGGATTACA-GCACGAGCCACTGTGCCCGGC 1229

 Qy 1380 catcattcctttttactgctgactaatagtctgctgtgtgaatccaccgctagaaaccca 1439
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1230 CATCATTCCTTTTTACTGCTGACTAATAGTCTGCTGTGTGAATCCACCGCTAGAAACCCA 1289

 Qy 1440 ctcatcagttgatggatcatgtgggttgccttctgctattcgcttattatgaacagtgtgg 1499
 ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
 Db 1290 CTCATCAGTTGATGGTCATGTGGGTGCTTCTCGTATTGCTTATTATGAACAGTGCTGG 1349

 Qy 1500 aataaacgttcctgtgcactcttgggcatacgccctaggagtggaaactgctgggt 1553
 ||||||||||||||| ||||||| ||||||||||||||||||||||||
 Db 1350 AATAAACGTTCTGTGCACTCTT-GGCATATGCCTAGGAGTGGAAGTGTGGGT 1402

Best hit oligo Seq ID 2

RESULT 2
 HSU14383
 LOCUS HSU14383 1403 bp mRNA linear PRI 31-DEC-1994 102(b)
 DEFINITION Human mucin (MUC8) mRNA, partial cds.
 ACCESSION U14383 U04799
 VERSION U14383.1 GI:606953
 KEYWORDS .
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 907 to 1403)
 AUTHORS Sachdev, G.P.
 TITLE Carboxy terminus of a major human tracheobronchial mucin MUC8
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 941)
 AUTHORS Shankar, V., Gilmore, M.S., Elkins, R.C. and Sachdev, G.P.
 TITLE A novel human airway mucin cDNA encodes a protein with unique
 tandem-repeat organization
 JOURNAL Biochem. J. 300 (Pt 2), 295-298 (1994)
 MEDLINE 94271137
 REFERENCE 3 (bases 1 to 1403)
 AUTHORS Sachdev, G.P.
 TITLE Direct Submission
 JOURNAL Submitted (03-SEP-1994) Goverdhan P. Sachdev, University of
 Oklahoma Health Sciences Center, Medicinal Chemistry, College of
 Pharmacy, 1110 N. Stonewall Avenue, Oklahoma City, OK 73117, USA
 COMMENT On Jan 1, 1995 this sequence version replaced gi:537430.
 FEATURES
 Location/Qualifiers
 source 1. .1403
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="12"
 /clone="PAM2"
 /tissue_type="trachea/bronchus"
 /dev_stage="adult"
 repeat_region <1. .892
 /rpt_type=tandem
 /rpt_unit=3. .43
 gene 1. .1403
 /gene="MUC8"
 CDS <1. .944
 /gene="MUC8"
 /codon_start=3
 /product="mucin"
 /protein_id="AAA58346.1"
 /db_xref="GI:501033"
 /translation="TSCPRPLQEGTPGSRAAHALSRRGHRVHELPTSSPGGDTGFMSC
 PRPFQEGTPGSRAAHVLSRKGPVHELPTSSPGRDPGFTSCPRPLQEGTRVTNCPRL
 QEGTPGSRAAHVLSRRGHRVHELPTSPGRDPGFMSCPRPLQEGTRVTNCPRLQEGT
 RVTSCPRRLQEGTRVTSCPRPLQEGTRVTNCPRALQEGTPGSRAAHALSRRGHRVHEL
 PTSSPGGDTGFTSCPRPLQEGTPGSRAAHALSRRGHRVHELPTSSPGRDPGHELPTSS
 PGGDTGFTSCPRTFQEGTPGSGLLPAAHIVPLCKSEER"
 3'UTR 945. .1403
 /gene="MUC8"
 polyA_signal 1350. .1355
 /gene="MUC8"

polyA_site 1403
/gene="MUC8"
/note="17 A residues"
BASE COUNT 254 a 490 c 413 g 246 t
ORIGIN

Query Match 21.1%; Score 334; DB 9; Length 1403;
Best Local Similarity 99.8%; Pred. No. 2.8e-181;
Matches 454; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
Qy 570 cacgaactgccacgcgctctccaggaggggacaccgggttcacgagctgccacgccct 629
    |||||
Db 566 CACGAAGTGCACGCGCTCTCCAGGAGGGGACACCGGGTTCACGAGCTGCCACGCCCT 625

Qy 630 ctccaggaagggaccccggttcacgagctgccacgtcctctccaggaggggacaccgg 689
    |||||
Db 626 CTCCAGGAAGGGACCCCGGGTTCACGAGCTGCCACGTCTCTCCAGGAGGGGACACCGG 685

Qy 690 gttcacgagctgccacgtcctctccaggaggggacaccgggttcacgagctgccacgc 749
    |||||
Db 686 GTTCACGAGCTGCCACGTCTCTCCAGGAGGGGACACCGGGTTCACGAGCTGCCACGC 745

Qy 750 cctctccaggaggggacaccgggttcacgagctgccacgtcctctccaggaagggaccc 809
    |||||
Db 746 CCTCTCCAGGAGGGGACACCGGGTTCACGAGCTGCCACGTCTCTCCAGGAAGGGACCC 805

Qy 810 gggtcacgagctgccacgtcctctccaggaggggacaccgggttcacgagctgccac 869
    |||||
Db 806 GGGT-CACGAGCTGCCACGTCTCTCCAGGAGGGGACACCGGGTTCACGAGCTGCCAC 864

Qy 870 gcactttccaggaagggaccccggttcaggtctcctgccggcccacatcgtgcctttgt 929
    |||||
Db 865 GCACTTTCAGGAAGGGACCCCGGGTTCAGGTCTCCTGCCGGCCCACATCGTGCCTTTGT 924

Qy 930 gtaaatcagaagaaagatgaggaacaggccctcctctctctccaggcaggctttggtgga 989
    |||||
Db 925 GTAAATCAGAAGAAAGATGAGGAACAGGCCCTCCTCTCTCTCCAGGCAGGCTTTGGTGGA 984

Qy 990 ggggctggatctcctgccgcaccttcctggcagg 1024
    |||||
Db 985 GGGGCTGGATCTCCTGCCGCACCTTCCCTGGCAGG 1019
```

encoding 3

LOCUS HSU14383 1403 bp mRNA linear PRI 31-DEC-1994
DEFINITION Human mucin (MUC8) mRNA, partial cds.
ACCESSION U14383 U04799
VERSION U14383.1 GI:606953
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 907 to 1403)
AUTHORS Sachdev, G.P.
TITLE Carboxy terminus of a major human tracheobronchial mucin MUC8
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 941)
AUTHORS Shankar, V., Gilmore, M.S., Elkins, R.C. and Sachdev, G.P.
TITLE A novel human airway mucin cDNA encodes a protein with unique
tandem-repeat organization
JOURNAL Biochem. J. 300 (Pt 2), 295-298 (1994)
MEDLINE 94271137
REFERENCE 3 (bases 1 to 1403)
AUTHORS Sachdev, G.P.
TITLE Direct Submission
JOURNAL Submitted (03-SEP-1994) Goverdhan P. Sachdev, University of
Oklahoma Health Sciences Center, Medicinal Chemistry, College of
Pharmacy, 1110 N. Stonewall Avenue, Oklahoma City, OK 73117, USA
COMMENT On Jan 1, 1995 this sequence version replaced gi:537430.
FEATURES
Location/Qualifiers
source 1. .1403
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="PAM2"
/tissue_type="trachea/bronchus"
/dev_stage="adult"
repeat_region <1. .892
/rpt_type=tandem
/rpt_unit=3. .43
gene 1. .1403
/gene="MUC8"
CDS <1. .944
/gene="MUC8"
/codon_start=3
/product="mucin"
/protein_id="AAA58346.1"
/db_xref="GI:501033"
/translation="TSCPRPLQEGTPGSRAAHALSRRGHRVHELPTSSPGGDTGFMSC
PRPFQEGTPGSRAAHVLSRKGPRVHELPTSSPGRDPGFTSCPRPLQEGTRVTNCPRL
QEGTPGSRAAHVLSRRGHRVHELPTSPGRDPGFMSCPRPLQEGTRVTNCPRLQEGT
RVTSCPRRLQEGTRVTSCPRPLQEGTRVTNCPRALQEGTPGSRAAHALSRRGPRVHEL
PTSSPGGDTGFTSCPRPLQEGTPGSRAAHALSRRGHRVHELPTSSPGRDPGHELPTSS
PGGDTGFTSCPRTFQEGTPGSGLLPAHIVPLCKSEER"
3'UTR 945. .1403
/gene="MUC8"
polyA_signal 1350. .1355
/gene="MUC8"
polyA_site 1403
/gene="MUC8"

498 CCACGTCGTCTCCAGGAAGGGACCCGGGTC.ACGAGCTGCCCACGTCCTC 546
 183 euGlnGluGlyThrArgValHisGluLeuProThrArgSerProGlyGly 199
 ||||||||||||||| |||||||||||||||||||||
 547 TCCAGGAAGGGACCCGGGT.CACGAACTGCCACGCGCTCTCCAGGAGGG 595
 200 AspThrGlyPheThrSerCysProArgProLeuGlnGluGlyThrProGl 216
 ||||||||||||||| |||||||||||||||||||||
 596 GACACCGGGTTCACGAGCTGCCACGCCCTCTCCAGGAAGGGACCCGGG 645
 216 ySerArgAlaAlaHisValLeuSerArgArgGlyHisArgValHisGluL 233
 ||||||||||||||| |||||||||||||||||||||
 646 TTCACGAGCTGCCACGTCCTCTCCAGGAGGGGACACCGGGTTCACGAGC 695
 233 euProThrSerSerProGlyGlyAspThrGlyPheThrSerCysProArg 249
 ||||||||||||||| |||||||||||||||||||||
 696 TGGCCACGTCCTCTCCAGGAGGGGACACCGGGTTCACGAGCTGCCACGC 745
 250 ProLeuGlnGluGlyThrProGlySerArgAlaAlaHisValLeuSerAr 266
 ||||||||||||||| |||||||||||||||||||||
 746 CCTCTCCAGGAGGGGACACCGGGTTCACGAGCTGCCACGTCCTCTCCAG 795
 266 gLysGlyProGlySerThrSerCysProArgProLeuGlnGluGlyThrP 283
 ||||||||||||||| |||||||||||||||||||||
 796 GAAGGGACCCGGGTC.ACGAGCTGCCACGTCCTCTCCAGGAGGGGACAC 844
 283 roGlySerArgAlaAlaHisAlaLeuSerArgLysGlyProArgValGln 299
 ||||||||||||||| |||||||||||||||||||||
 845 CGGGTTCACGAGCTGCCACGCACTTTCCAGGAAGGGACCCGGGTTCAG 894
 300 ValSerCysArgProThrSerCysLeuCysValAsnGlnLysLysAspGl 316
 ||||||||||||||| |||||||||||||||||||||
 895 GTCTCCTGCCGGCCACATCGTGCCTTTGTGTAAATCAGAAGAAAGATGA 944
 316 uGluGlnAlaLeuLeuSerLeuGlnAlaGlyPheGlyGlyGlyAlaGlyS 333
 ||||||||||||||| |||||||||||||||||||||
 945 GGAACAGGCCCTCTCTCTCTCCAGGCAGGCTTTGGTGGAGGGGCTGGAT 994
 333 erProAlaAlaProSerLeuAlaGlyHisProValLeuGluProGlnAsn 349
 ||||||||||||||| |||||||||:::|||||||
 995 CTCCTGCCGCACCTTCCCTGGCAGG.CACCCTGTCGTTGAGCCCCAGAAC 1043
 350 CysArgArgProAlaGluLysGlySerMetMetAlaProArgCys.AlaA 366
 ||||||||||||||| |||||
 1044 TGCAGGCGGCCCGGCAGAGAAGGGGTCCATGATGGCGCCTCGGTGCGGCC. 1092
 366 laLeuAspLeuProProTrpThrTrpGluProProGlySerSerHisSer 382
 |||||||||||||||
 1093 ..TTGGACCTGCCCCCATGGACCTGG..... 1116
 383 GlyLysGluGlySerGlyHisGlyGlyArgProGlyProIleProValPr 399
 1116 1116
 399 oTrpProPhePheLeuLeuProValCysHisCysProGlyAlaPheAlaP 416
 1116 1116

416 roAlaPheProLeuSerArgGlnGlyPheSerSerLeuAlaArgLeuVal 432
||||||||||||||||||||||||||||||
1117AGACAGGGTTTCTCCTCATTGGCCAGGCTGGTC 1149

433 SerAsnSer 435
|||||||
1150 TCGAACTCC 1158

seq_name: gb_vi:U97553

102 (a or b)
SID 48 entire

RESULT 1
AI126846
LOCUS AI126846 464 bp mRNA EST 26-OCT-1998
DEFINITION qb86d03.x1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
IMAGE:1706981 3' similar to contains Alu repetitive element
;contains L1.t1 L1 L1 repetitive element ;, mRNA sequence.
ACCESSION AI126846
VERSION AI126846.1 GI:3595360
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 464)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 437 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 421.
FEATURES Location/Qualifiers
source 1. .464
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1706981"
/clone_lib="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbHL19W."
BASE COUNT 113 a 116 c 95 g 140 t
ORIGIN

Query Match 100.0%; Score 39; DB 10; Length 464;
Best Local Similarity 100.0%; Pred. No. 3.2e-05;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tgtgcactcttgggcatacgccctaggagtgggaactgctg 39
|||||
Db 271 TGTGCACTCTTGGGCATACGCCTAGGAGTGGAACTGCTG 309

SID 47
oligo
102(e)

RESULT 2
US-09-058-489-38/c
; Sequence 38, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahn, Bruce
; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; TITLE OF INVENTION: the Y Chromosome
; FILE REFERENCE: WHI97-08pA
; CURRENT APPLICATION NUMBER: US/09/058,489
; CURRENT FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/041,877
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 38
; LENGTH: 1964
; TYPE: DNA
; ORGANISM: Human
US-09-058-489-38

Query Match 40.0%; Score 16; DB 3; Length 1964;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 accaccgtgtcctgct 29
|||||
Db 852 ACCACCGTGTCTGCT 837

oligo sid
AS

don't have SNP

RESULT 6
G26698
LOCUS G26698 462 bp DNA linear STS 02-JUN-1996
DEFINITION human STS TH, sequence tagged site.
ACCESSION G26698
VERSION G26698.1 GI:1348930
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE human STSs derived from sequences in dbEST and the Unigene collection.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 462)
AUTHORS Hudson, T.
TITLE Whitehead Institute/MIT Center for Genome Research; Physically Mapped STSs
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 462)
AUTHORS Hudson, T.
TITLE Whitehead Institute/MIT Center for Genome Research; Physically Mapped ESTs
JOURNAL Unpublished
COMMENT
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu

Primer A: GAACAGGCCCTCCTCTCTCT
Primer B: ATGGACCCCTTCTCTGCC
STS size: 127
PCR Profile:
Presoak:
Denaturation:
Annealing: 56 degrees C
Polymerization:
PCR Cycles: 35
Thermal Cycler:
Protocol:
Template: 10 ng
Primer: each 5 pM
dNTPs: each 4 nM
Taq Polymerase: 0.025 units/ul
Total Vol: 20 ul

Buffer:
MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCL: 10 mM
pH: 9.3

Derived from dbEST (genbank accession U14383).
FEATURES
source 1..462

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="364_G_6"
STS      5. .131
primer_bind 5. .24
primer_bind complement(114. .131)
BASE COUNT      66 a      107 c      104 g      91 t      94 others
ORIGIN

```

```

Query Match          51.3%; Score 20; DB 11; Length 462;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 20; Conservative . 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      20 gcctaggagtgggaactgctg 39
        |||
Db      439 GCCTAGGAGTGGAAGTCTG 458

```

0190
A6

RESULT 1
AA463832/c

LOCUS AA463832 206 bp mRNA linear EST 10-JUN-1997

DEFINITION zx67e07.r1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
IMAGE:796548 5' similar to contains Alu repetitive element;contains
element PTR7 repetitive element ;; mRNA sequence.

ACCESSION AA463832

VERSION AA463832.1 GI:2188716

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 206)

AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.,
Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,
T., Waterston,R. and Wilson,R.

TITLE WashU-Merck EST Project 1997

JOURNAL Unpublished (1997)

COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Putative full length read
The vector to vector length is 207
Seq primer: -28ml3 rev2 ET from Amersham.

FEATURES

source 1. .206
/organism="Homo sapiens"
/db_xref="GDB:6040837"
/db_xref="taxon:9606"
/clone="IMAGE:796548"
/clone_lib="Soares_total_fetus_Nb2HF8_9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGGCCGCTTAATTTTTTTTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 59 a 45 c 61 g 41 t

ORIGIN

Query Match 65.0%; Score 26; DB 9; Length 206;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 tgcaccaccatgtcctgctaattttt 36
 |||
 Db 76 TGCACCACCATGTCTGCTAATTTTT 51

SIP 47
102(b) oligo

RESULT 4
HSAC000367
LOCUS HSAC000367 43349 bp DNA linear PRI 28-AUG-1997
DEFINITION Human Cosmid gl862x055 from 7q31.3, complete sequence.
ACCESSION AC000367
VERSION AC000367.1 GI:2347066
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 43349)
AUTHORS Iadonato,S.P., Yu,J., Wong,G.K.-S., Magness,C.L., Green,E.D.,
Green,P. and Olson,M.V.
TITLE Large-scale MCD Mapping and Sequencing of Human Chromosome 7
JOURNAL Unpublished (1996)
REFERENCE 2 (bases 1 to 43349)
AUTHORS Iadonato,S.P., Yu,J., Wong,G.K.-S., Magness,C.L., Green,E.D.,
Green,P. and Olson,M.V.
TITLE Direct Submission
JOURNAL Submitted (21-OCT-1996) Human Genome Center, University of
Washington, Box 352145, Seattle, WA 98195, USA
REMARK University of Washington Human Genome Center
Box 352145 Seattle, WA 98195
Contact: Shawn Iadonato (iadonato@u.washington.edu)
Large-scale MCD Mapping and Sequencing of Human Chromosome 7
REFERENCE 3 (bases 1 to 43349)
AUTHORS Magness,C.L.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-1997) Human Genome Center, University of
Washington, Box 352145, Seattle, WA 98195, USA
REFERENCE 4 (bases 1 to 43349)
AUTHORS Iadonato,S.P., Yu,J., Wong,G.K.-S., Magness,C.L., Green,E.D.,
Green,P. and Olson,M.V.
TITLE Direct Submission
JOURNAL Submitted (28-AUG-1997) Human Genome Center, University of
Washington, Box 352145, Seattle, WA 98195, USA
REMARK University of Washington Human Genome Center
Box 352145 Seattle, WA 98195
Contact: Shawn Iadonato (iadonato@u.washington.edu)
COMMENT On Aug 28, 1997 this sequence version replaced gi:1881556.
Overlapping Sequences:
5': UWGC:gl862x083 (Genbank Accession: AC002113)
3': UWGC:gl862d218 (Genbank Accession: AC000373)

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Pct of Consensus above quality 40: 98.4%
Number of manually edited bases: 6

Double stranded (DS) coverage: 88.0%
DS or two chemistry coverage: 99.2%

Single stranded regions:

2

Quality Cumulative Percentage of Consensus

```

-----
90  xx (11.5%)
80+ xxxxxxxxxx (50.4%)
70+ xxxxxxxxxxxxxxxx (82.3%)
60+ xxxxxxxxxxxxxxxx (91.4%)
50+ xxxxxxxxxxxxxxxx (95.8%)
40+ xxxxxxxxxxxxxxxx (98.4%)
30+ xxxxxxxxxxxxxxxx (99.8%)
20+ xxxxxxxxxxxxxxxx (99.9%)
10+ xxxxxxxxxxxxxxxx (100.0%)
00+ xxxxxxxxxxxxxxxx (100.0%)

```

Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest Mapping. Comparison of the experimentally derived map digest fragments with sequence-predicted fragments is given below. Small fragments below a variable cutoff (approximately 400-600bp) are not mapped and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragment groups are separated by dashed lines.

EcoRI		HindIII		NsiI	
Map	Seq	Map	Seq	Map	Seq
5124.86	5144.00	4213.20	4271.00	2634.83	2642.00
1562.97	1593.00	7922.52	7935.00	1789.10	1797.00
5177.32	5193.00	490.78	485.00	699.88	702.00
3960.50	3986.00	3168.35	3198.00	5150.11	5177.00
9380.91	9346.00	2464.68	2438.00	1216.04	1206.00
5538.62	5591.00	6305.33	6300.00	2719.74	2726.00
2565.93	2552.00	2757.13	2756.00	1428.27	1446.00
4417.36	4392.00	6579.15	6623.00	5504.09	5500.00
1651.21	1636.00	5153.00	5085.00	3214.21	3189.00
				1088.29	1089.00
				4175.13	4192.00
				2292.35	2307.00
				820.75	817.00

1471.36	1489.00
-----	-----
1022.53	1021.00
-----	-----
1338.53	1334.00
-----	-----
1602.67	1386.00
-----	-----
2084.45	2104.00
-----	-----

FEATURES	Location/Qualifiers
source	1. .43349 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="7" /map="7q31.3" /clone="NHGRI:yWSS1862" /sub_clone="UWGC:g1862x055" /cell_line="GM10791" /clone_lib="E. Green Chromosome 7 YAC Resource"
repeat_region	complement(688. .2990) /rpt_family="Tigger1"
repeat_region	complement(9460. .9602) /rpt_family="MER3"
repeat_region	complement(10588. .10883) /rpt_family="ALU"
repeat_region	10949. .11278 /rpt_family="ALU"
repeat_region	complement(22058. .22282) /rpt_family="MER3"
repeat_region	complement(24214. .25023) /rpt_family="L1"
repeat_region	25021. .25070 /rpt_family="L1"
repeat_region	complement(30836. .31068) /rpt_family="L1"
repeat_region	complement(32584. .32877) /rpt_family="ALU"
repeat_region	complement(38416. .38541) /rpt_family="ALU"

BASE COUNT 13710 a 8483 c 7675 g 13481 t
ORIGIN

Query Match 50.0%; Score 20; DB 9; Length 43349;
Best Local Similarity 100.0%; Pred. No. 0.89;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 tacagggtttgcaccaccgtg 22
 |||||
Db 32851 TACAGGTTTGCACCACCGTG 32870

plz SID 48 *don't have SNP*

```

RESULT 7
HSU14383
LOCUS      HSU14383                1403 bp    mRNA    linear    PRI 31-DEC-1994
DEFINITION Human mucin (MUC8) mRNA, partial cds.
ACCESSION  U14383 U04799
VERSION    U14383.1  GI:606953
KEYWORDS   .
SOURCE     human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 907 to 1403)
  AUTHORS  Sachdev,G.P.
  TITLE    Carboxy terminus of a major human tracheobronchial mucin MUC8
  JOURNAL  Unpublished
REFERENCE  2 (bases 1 to 941)
  AUTHORS  Shankar,V., Gilmore,M.S., Elkins,R.C. and Sachdev,G.P.
  TITLE    A novel human airway mucin cDNA encodes a protein with unique
            tandem-repeat organization
  JOURNAL  Biochem. J. 300 (Pt 2), 295-298 (1994)
  MEDLINE  94271137
REFERENCE  3 (bases 1 to 1403)
  AUTHORS  Sachdev,G.P.
  TITLE    Direct Submission
  JOURNAL  Submitted (03-SEP-1994) Goverdhan P. Sachdev, University of
            Oklahoma Health Sciences Center, Medicinal Chemistry, College of
            Pharmacy, 1110 N. Stonewall Avenue, Oklahoma City, OK 73117, USA
COMMENT    On Jan 1, 1995 this sequence version replaced gi:537430.
FEATURES   Location/Qualifiers
    source          1. .1403
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="12"
                    /clone="PAM2"
                    /tissue_type="trachea/bronchus"
                    /dev_stage="adult"
    repeat_region   <1. .892
                    /rpt_type=tandem
                    /rpt_unit=3. .43
    gene            1. .1403
                    /gene="MUC8"
    CDS             <1. .944
                    /gene="MUC8"
                    /codon_start=3
                    /product="mucin"
                    /protein_id="AAA58346.1"
                    /db_xref="GI:501033"
                    /translation="TSCPRPLQEGTPGSRAAAHALSRGHRVHELPTSSPGGDTGFMSC
PRPFQEGTPGSRAAHVLSRKGPVHELPTSSPGRDPGFTSCPRPLQEGTRVTNCPRPL
QEGTPGSRAAHVLSRRGHRVHELPTSPGRDPGFMSCPRPLQEGTRVTNCPRPLQEGT
RVTSCPRRLQEGTRVTSCPRPLQEGTRVTNCPRALQEGTPGSRAAAHALSRKGPRVEL
PTSSPGGDTGFTSCPRPLQEGTPGSRAAAHALSRGHRVHELPTSSPGRDPGHELPTSS
PGGDTGFTSCPRTFQEGTPGSGLLP AHIVPLCKSEER"
    3'UTR          945. .1403
                    /gene="MUC8"
    polyA_signal    1350. .1355
                    /gene="MUC8"

```

polyA_site 1403
/gene="MUC8"
/note="17 A residues"
BASE COUNT 254 a 490 c 413 g 246 t
ORIGIN

Query Match 51.3%; Score 20; DB 9; Length 1403;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 gcctaggagtgggaactgctg 39
|||||||
Db 1380 GCCTAGGAGTGGAAGTCTG 1399

102(a or b)
origo 50

RESULT 5
SCE63
LOCUS SCE63 37200 bp DNA linear BCT 17-MAR-1999
DEFINITION Streptomyces coelicolor cosmid E63.
ACCESSION AL035640
VERSION AL035640.2 GI:4500374
KEYWORDS 4-hydroxyphenylpyruvic acid dioxygenase; aminotransferase;
AMP-binding domain; CDA peptide synthetase; cdaPS1; cdaPS2; cdaPS3;
DUF4, Domain of unknown function; glycolate oxidase;
Phosphopantetheine attachment site.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Streptomyces coelicolor A3(2)
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 37200)
AUTHORS Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
Kinashi,H. and Hopwood,D.A.
TITLE A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)
MEDLINE 97000351
REFERENCE 2 (bases 1 to 37200)
AUTHORS Saunder,D.C. and Harris,D.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 37200)
AUTHORS Bentley,S.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
TITLE Direct Submission
JOURNAL Submitted (17-MAR-1999) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT On Mar 24, 1999 this sequence version replaced gi:4481931.
Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC.
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL; http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are
numbered using the following system eg SC7B7.01c. SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length
in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon. Gene
prediction is based on positional base preference in codons using a
specially developed Hidden Markov Model (Krogh et al., Nucleic
Acids Research, 22(22):4768-4778(1994)) and the FramePlot program
of Bibb et al., Gene 30:157-66(1984) as implemented at
<http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
codon (atg, gtg, ttg or (att)) which is preceded by an upstream

ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid E63 lies between E8 and E29 on the AseI-E genomic restriction fragment.

FEATURES	Location/Qualifiers
source	1. .37200 /organism="Streptomyces coelicolor A3(2)" /strain="A3(2)" /db_xref="taxon:100226" /clone="cosmid E63"
gene	complement(1. .138) /gene="SCE63.06"
CDS	complement(<1. .138) /gene="SCE63.06" /note="SCE63.06, partial CDS, possible aminotransferase, len: >45aa; similar to the N-terminal region of TR:052815 (EMBL:AJ223998) a protein similar to aminotransferase from a cluster of genes nvolved in the biosynthesis of a vancomycin group antibiotic in Amycolatopsis orientalis (Actinomycete) (438 aa) fasta scores; opt: 87, z-score: 147.0, E(): 0.69, (45.2% identity in 31 aa overlap). The remainder of this CDS lies on the neighbouring Streptomyces coelicolor cosmid E8." /codon_start=1 /transl_table=11 /label=SCE63.06 /product="putative aminotransferase" /protein_id="CAB38521.1" /db_xref="GI:4481937" /translation="MTTTTPARTDRDGVLTARTALHPSLSAPILDTMFLNEVTLRYFQAI"
gene	complement(135. .1268) /gene="SCE63.05"
CDS	complement(135. .1268) /gene="SCE63.05" /note="SCE63.05, probable glycolate oxidase, len: 377aa; similar to many both prokaryotic and eukaryotic egs. TR:052792 (EMBL:AJ223998) similar to glycolate oxidase from a cluster of genes nvolved in the biosynthesis of a vancomycin group antibiotic in Amycolatopsis orientalis (Actinomycete) (357 aa) fasta scores; opt: 995, z-score: 998.4, E(): 0, (47.5% identity in 343 aa overlap) and SW:GOX_SPIOL glycolate oxidase from Spinacia oleracea (spinach) (369 aa) fasta scores; opt: 917, z-score: 920.7, E(): 0, (41.9% identity in 360 aa overlap). Contains Pfam match to entry PF01070 FMN_dh, FMN-dependent dehydrogenase, score 412.60, E-value 3.6e-120 and PS00557 FMN-dependent alpha-hydroxy acid dehydrogenases active site." /codon_start=1 /transl_table=11 /label=SCE63.05 /product="putative glycolate oxidase"

```

/protein_id="CAB38520.1"
/db_xref="GI:4481936"
/translation="MREPLTLDDFARLARGQLPAATWDFIAGGAGRERTLAANEAVFG
AVRLRPALPGIEEPDTSVEVLGSRWPAPVGIAPVAYHGLAHFDGEPATAAAAGALGL
PLVVSTFAGRSLEEVARAASAPLWLQLYCFRDHETTLGLARRARDSGYQALVLTVDTP
FTGRRRLDLRNGFAVPAHITPANLTGTAAAGSATPGAHSRLAFDRRLDWSFVARLGAA
SGLPVLAAGVLTAPDAEAAVAAGVAGIVVSNHGGRLDGAATLEALPEVVS AVRGR
PVLDDGGVRTGADVLAALALGARAVLVGRPALYALAVGGASGVRRLTLLTEDFADTM
VLTGHAATGTIGPDTLAPPHHAPPHHGPPPTAPRPAPHRDRSHG"
misc_feature complement(213..1229)
/gene="SCE63.05"
/note="Pfam match to entry PF01070 FMN_dh, FMN-dependent
de hydrogenase, score 412.60, E-value 3.6e-120"
misc_feature complement(507..527)
/gene="SCE63.05"
/note="PS00557 FMN-dependent alpha-hydroxy acid
dehydrogenase active site."
/label="*"
gene complement(1418..2533)
/gene="SCE63.04"
CDS complement(1418..2533)
/gene="SCE63.04"
/note="SCE63.04, probable 4-hydroxyphenylpyruvic acid
dioxygenase, len: 371aa; similar to many both prokaryotic
and eukaryotic egs. TR:O52791 (EMBL:AJ223998) similar to
hydroxyphenyl pyruvate dioxygenase from a cluster of genes
involved in the biosynthesis of a vancomycin group
antibiotic in Amycolatopsis orientalis (Actinomycete) (357
aa) fasta scores; opt: 989, z-score: 1130.3, E(): 0,
(48.7% identity in 355 aa overlap) and SW:HPPD_MOUSE hpd,
4-hydroxyphenylpyruvic acid dioxygenase from Mus musculus
(mouse) (392 aa) fasta scores; opt: 610, z-score: 698.5,
E(): 1.3e-31, (31.6% identity in 361 aa overlap)."
/codon_start=1
/transl_table=11
/label="SCE63.04"
/product="putative 4-hydroxyphenylpyruvic acid
dioxygenase"
/protein_id="CAB38519.1"
/db_xref="GI:4481935"
/translation="MLPPFPFLHWRAAMPSPDIAYAELYVALDREASGFLVDSLGFVP
LAVAGPATGTHDRRSTVLRSGEVTLVVTQALAPDTPVARYVERHGDSIADLAFGCDDV
RSCFDRAVLAGEALQAPTPSHRAGQDAWFATVSGFGDIRHTLVPAADGDGAGLLPPD
RDWALLPAATGRTGPRPLLDHVAVCLESGTLRSTAEFYEAADFMPYYSSEYIEVGEQA
MDSIVVRNAGGGITFTLIEPDDTRVPGQIDQFLSAHDGPGVQHFLAVLDDIVGSVRSI
GDRGVAFLRTPGAYYDLLTERVGAMADAIEDLRETNVLADRDEWGYLLQIFTRSPYPR
GTLFYEYIQRNGARGFGSSNIKALYEAVEREVAGR"
gene 2802..25193
/gene="SCE63.03c"
/note="cdaPSI"
CDS 2802..25193
/gene="SCE63.03c"
/note="SCE63.03c, cdaPSI, CDA peptide synthetase I, len:
7463aa; part of the calcium-dependent antibiotic (CDA)
biosynthetic cluster from Streptomyces coelicolor. CDA is
a peptide antibiotic which is synthesised non-ribosomally
by a putative multifunctional peptide synthetase enzyme.

```

This CDS encodes a subunit of this enzyme and is suspected to be responsible for the incorporation of the first 6 amino acids into the antibiotic structure. This ORF overlaps the downstream (cdaPSII) ORF by one base indicating possible translational coupling. Contains eight separate Pfam matches to entry PF00668 DUF4, Domain of unknown function (U), six separate Pfam matches to entry PF00501 AMP-binding, AMP-binding enzyme (A) and six separate Pfam matches to entry PF00550 pp-binding, Phosphopantetheine attachment site (P). These Pfam matches cover the full length of the protein in the following order from N to C-terminal

U-A-P-U-A-P-U-A-P-U-U-A-P-U-A-P-U-A-P-U."

/codon_start=1

/transl_table=11

/label=cdaPSI

/product="CDA peptide synthetase I"

/protein_id="CAB38518.1"

/db_xref="GI:4481934"

/translation="MSENSSVRHGLTSAQHEVWLAQQLDPRGAHYRTGSCLEIDGPLD
HAVLSRALRLTVAGTETLCSRFLTDEEGRPYRAYCPPAPEGSAAVEDPDGVPYTPVLL
RHIDLSGHEDPEGEAQRWMDRDRATPLPLDRPGLSSHALFTIGGGRHLYYLGVHHIVI
DGTSMALFYERLAEVYRALRDGRAVPAAAFGDTDRMVAGEEAYRASARYERDRAYWTG
LFTDRPEPVSLTGRGGGRALAPTVRSLGLPPERTEVLGRAAEATGAHWARVVIAGVAA
FLHRTTGARDVVVSVVPTGRYGANARITPGMVSNRLPLRLAVRPGESFARVVETVSEA
MSGLLAHSRFRGEDLDRELGGAGVSGPTVNVMPYIRPVDFGC-PVGLMRSISSGPTTDL
NIVLTGTPESGLRVDFEGNPQVYGGQDLTVLQERFVRFLAEIAADPAATVDEVALLTP
DERERVLDGWNDTAHEVPETTLPELFAARAARTPGHEALVYEGTSLTYAELDARAERL
AGALTARGAGPERFVAVAVERSAELVVALLAVLKSGAAYVPVDPGYPADRIAHILRDA
GAMLVLTTRDTAERLPGDGTPrLLLDPEAAAGTTAAGAPAPFGTLPRALPAPGHPAYV
IYTSGSTGRPKGVVISHRAIVNRLAWMQDTYGLEPSDRVLQKTPSGFDVSVWEFFWPL

Query Match 43.6%; Score 17; DB 1; Length 37200;

Best Local Similarity 100.0%; Pred. No. 25;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 ctgcgccacctaacc 22

|||||||

Db 13854 CTGCGCCACCTCAACCC 13870

oligo 48 102(e) has SNP

RESULT 4
US-09-103-330-35
; Sequence 35, Application US/09103330A
; Patent No. 6319716
; GENERAL INFORMATION:
; APPLICANT: TIKOO, SURESH K.
; APPLICANT: BABIUK, LORNE A.
; APPLICANT: REDDY, POLICE S.
; TITLE OF INVENTION: ISOLATION OF MUTANTS IN THE E3 REGION OF THE
; TITLE OF INVENTION: BOVINE ADENOVIRUS GENOME AND THEIR USE IN VACCINES
; FILE REFERENCE: 293102002121
; CURRENT APPLICATION NUMBER: US/09/103,330A
; CURRENT FILING DATE: 1998-06-23
; EARLIER APPLICATION NUMBER: 08/880,234
; EARLIER FILING DATE: 1997-06-23
; EARLIER APPLICATION NUMBER: 08/164,292
; EARLIER FILING DATE: 1993-12-09
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35
; LENGTH: 34446
; TYPE: DNA
; ORGANISM: Bovine adenovirus type 3
US-09-103-330-35

Query Match 38.5%; Score 15; DB 4; Length 34446;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ctcttgggcatacgc 21
|||||||
Db 13657 ctcttgggcatacgc 13671

list 49

RESULT 1
G26698
LOCUS G26698 462 bp DNA linear STS 02-JUN-1996
DEFINITION human STS TH, sequence tagged site.
ACCESSION G26698
VERSION G26698.1 GI:1348930
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE human STSs derived from sequences in dbEST and the Unigene collection.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 462)
AUTHORS Hudson,T.
TITLE Whitehead Institute/MIT Center for Genome Research; Physically Mapped STSs

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 462)

AUTHORS Hudson,T.
TITLE Whitehead Institute/MIT Center for Genome Research; Physically Mapped ESTs

JOURNAL Unpublished

COMMENT

Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu

Primer A: GAACAGGCCCTCCTCTCTCT

Primer B: ATGGACCCCTTCTCTGCC

STS size: 127

PCR Profile:

Presoak:

Denaturation:

Annealing: 56 degrees C

Polymerization:

PCR Cycles: 35

Thermal Cycler:

Protocol:

Template: 10 ng

Primer: each 5 pM

dNTPs: each 4 nM

Taq Polymerase: 0.025 units/ul

Total Vol: 20 ul

Buffer:

MgCl2: 1.5 mM

KCl: 50 mM

Tris-HCL: 10 mM

pH: 9.3

Derived from dbEST (genbank accession U14383).

FEATURES Location/Qualifiers

source 1..462

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="364_G_6"
STS      5. .131
primer_bind 5. .24
primer_bind complement(114. .131)
BASE COUNT      66 a      107 c      104 g      91 t      94 others
ORIGIN

```

```

Query Match      69.2%; Score 27; DB 11; Length 462;
Best Local Similarity 100.0%; Pred. No. 2.2e-05;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      13 ggcataatgcctaggagtggaaactgctg 39
        |||
Db      432 GGCATATGCCTAGGAGTGGAAC TGCTG 458

```

** used this for 49*

```

RESULT 2
HSU14383
LOCUS      HSU14383      1403 bp      mRNA      linear      PRI 31-DEC-1994
DEFINITION Human mucin (MUC8) mRNA, partial cds.
ACCESSION  U14383 U04799
VERSION    U14383.1 GI:606953
KEYWORDS   .
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 907 to 1403)
AUTHORS    Sachdev,G.P.
TITLE      Carboxy terminus of a major human tracheobronchial mucin MUC8
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 941)
AUTHORS     Shankar,V., Gilmore,M.S., Elkins,R.C. and Sachdev,G.P.
TITLE      A novel human airway mucin cDNA encodes a protein with unique
            tandem-repeat organization
JOURNAL    Biochem. J. 300 (Pt 2), 295-298 (1994)
MEDLINE    94271137
REFERENCE  3 (bases 1 to 1403)
AUTHORS     Sachdev,G.P.
TITLE      Direct Submission
JOURNAL    Submitted (03-SEP-1994) Goverdhan P. Sachdev, University of
            Oklahoma Health Sciences Center, Medicinal Chemistry, College of
            Pharmacy, 1110 N. Stonewall Avenue, Oklahoma City, OK 73117, USA
COMMENT    On Jan 1, 1995 this sequence version replaced gi:557430.
FEATURES   Location/Qualifiers
            source      1. .1403
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="12"
                        /clone="PAM2"
                        /tissue_type="trachea/bronchus"
                        /dev_stage="adult"
            repeat_region <1. .892
                        /rpt_type=tandem

```

```

gene          /rpt_unit=3. .43
              1. .1403
              /gene="MUC8"
CDS           <1. .944
              /gene="MUC8"
              /codon_start=3
              /product="mucin"
              /protein_id="AAA58346.1"
              /db_xref="GI:501033"
              /translation="TSCPRPLQEGTPGSRAAHALSRRGHRVHELPTSSPGGDTGFMSC
PRPFQEGTPGSRAAHVLSRKGPRVHELPTSSPGRDPGFTSCPRPLQEGTRVTNCPRPL
QEGTPGSRAAHVLSRRGHRVHELPTSPGRDPGFMSCPRPLQEGTRVTNCPRPLQEGT
RVTSCPRRLQEGTRVTSCPRPLQEGTRVTNCPRALQEGTPGSRAAHALSRRGHRVHEL
PTSSPGGDTGFTSCPRPLQEGTPGSRAAHALSRRGHRVHELPTSSPGRDPGHELPTSS
PGGDTGFTSCPRTFQEGTPGSGLLPAHIVPLCKSEER"
3'UTR        945. .1403
              /gene="MUC8"
polyA_signal  1350. .1355
              /gene="MUC8"
polyA_site    1403
              /gene="MUC8"
              /note="17 A residues"
BASE COUNT   254 a    490 c    413 g    246 t
ORIGIN

```

```

Query Match          69.2%; Score 27; DB 9; Length 1403;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      13 ggcataatgcctaggagtggaaactgctg 39
          |||
Db      1373 GGCATATGCCTAGGAGTGGAAGTCTG 1399

```

RESULT 11

AAV37951

ID AAV37951 standard; DNA; 2962 BP.

XX

AC AAV37951;

XX

DT 10-SEP-1998 (first entry)

XX

DE Human erythropoietin clone 7.2 encoding DNA.

XX

KW Human; erythropoietin; EPO; Chinese hamster ovary cell; CHO; strain;

KW medicine; biological research; ss.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 625..2772

FT /*tag= a

FT /product= "erythropoietin"

FT /note= "contains introns"

FT exon 625..640

FT /*tag= b

FT /number= 1

FT intron 641..1201

FT /*tag= c

FT /number= 1

FT exon 1202..1347

FT /*tag= d

FT /number= 2

FT intron 1348..1605

FT /*tag= e

FT /number= 2

FT exon 1606..1692

FT /*tag= f

FT /number= 3

FT intron 1693..2304

FT /*tag= g

FT /number= 3

FT exon 2305..2484

FT /*tag= h

FT /number= 4

FT intron 2485..2616

FT /*tag= i

FT /number= 4

FT exon 2617..2772

FT /*tag= j

FT /number= 5

XX

PN RU2089611-C1.

XX

PD 10-SEP-1997.

XX

PF 13-JUL-1995; 95RU-0111858.

XX

PR 13-JUL-1995; 95RU-0111858.

XX

PA (MEDB=) MED BIOTECHN RES PRODN CENTRE.

102 b
slip 50

XX
 PI Kamerova IA, Kolobkov SL, Zelenin MG;
 XX
 DR WPI; 1998-205757/18.
 DR P-PSDB; AAW62048.
 XX
 PT New strain of cultivated cells of Chinese hamster - acts as producer
 PT of human erythropoietin which can be used in medicine and in
 PT biological research
 XX
 PS Disclosure; Column 15-22; 13pp; English.
 XX
 CC The present sequence encodes human erythropoietin clone 7.2 from
 CC the present invention. The present invention describes a new CHO
 CC strain of cultivated cells of Chinese hamster VSKK (P) 637 D, which
 CC produces human erythropoietin. The new strain is used as a new
 CC strain-producer of human erythropoietin, which can be used in medical
 CC therapy and research, and also in biological research. The use of the
 CC strain reduces the cost of production of human erythropoietin owing
 CC to increased productivity of the strain.
 XX
 SQ Sequence 2962 BP; 596 A; 909 C; 881 G; 576 T; 0 other;

Query Match 38.5%; Score 15; DB 19; Length 2962;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 cctcaacccaggcgt 28
 |||||
 Db 248 cctcaacccaggcgt 262

RESULT 12

AAA71992

ID AAA71992 standard; DNA; 2962 BP.

XX

AC AAA71992;

XX

DT 19-JAN-2001 (first entry)

XX

DE Human erythropoietin DNA from clone 7.2.

XX

KW Erythropoietin; human; antianemic; late erythrocyte precursor cell;
 KW replacement therapy; treatment; ds.

XX

OS Homo sapiens.

XX

FH	Key	Location/Qualifiers
FT	CDS	625..2772
FT		/*tag= a
FT		/product= "erythropoietin"
FT	sig_peptide	625..1269
FT		/*tag= b
FT	mat_peptide	1270..2769
FT		/*tag= c
FT	exon	625..640

```

FT          /*tag= d
FT          /number= 1
FT  intron  641..1201
FT          /*tag= e
FT          /number= 1
FT  exon    1202..1347
FT          /*tag= f
FT          /number= 2
FT  intron  1348..1605
FT          /*tag= g
FT          /number= 2
FT  exon    1606..1693
FT          /*tag= h
FT          /number= 3
FT  intron  1694..2304
FT          /*tag= i
FT          /number= 3
FT  exon    2305..2484
FT          /*tag= j
FT          /number= 4
FT  intron  2485..2616
FT          /*tag= k
FT          /number= 4
FT  exon    2617..2772
FT          /*tag= l
FT          /number= 5

```

XX

PN DE19855489-A1.

XX

PD 17-AUG-2000.

XX

PF 01-DEC-1998; 98DE-1055489.

XX

PR 01-DEC-1998; 98DE-1055489.

XX

PA (GROZ/) GROZA I.

XX

DR WPI; 2000-566040/53.

DR P-PSDB; AAB10654.

XX

PT New nucleic acid molecule comprising simian virus 40 regulatory
PT sequences and antibiotic resistance gene, useful for expressing
PT erythropoietin in mammalian cells for treating anemia -

XX

PS Claim 1; Fig 5; 18pp; German.

XX

CC This invention describes a novel nucleic acid molecule (I) encoding an
CC erythropoietin (EPO) polypeptide (II), transcriptional and translational
CC regulatory sequences from simian virus 40 (SV40), including the SV40
CC early promoter and a sequence encoding resistance to an antibiotic. The
CC product of the invention has antianemic activity. EPO regulates
CC proliferation and differentiation of late erythrocyte precursor cells.
CC (I) is used for the recombinant production of human EPO in mammalian
CC cells. EPO is used, in replacement therapy, to treat anemia. Cells
CC transformed with (I) produce EPO at a high level (e.g. 1500-1800
CC international units/ml) which is stable under non-selection conditions.
CC The plasmid copy number in the cells can be increased without using the

CC expensive and highly cytostatic agent methotrexate. This sequence
CC encodes the human erythropoietin protein which is described in the method
CC of the invention.
XX
SQ Sequence 2962 BP; 591 A; 895 C; 872 G; 573 T; 31 other;

Query Match 38.5%; Score 15; DB 21; Length 2962;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 cctcaacccaggcgt 28
|||||
Db 248 cctcaacccaggcgt 262

RESULT 13
AAN50347
ID AAN50347 standard; DNA; 3211 BP.
XX
AC AAN50347;
XX
DT 01-JAN-1980 (first entry)
XX
DE Positive clone (phage lambda-hE1) isolated from human fetal liver
DE gene bank.
XX
KW Erythropoietin; red blood cell; erythrocyte; anaemia; blood;
KW disorder; ss; phage lambda-hE1; gene bank.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT intron 621..632
FT /*tag= a
FT intron 1072..1218
FT /*tag= b
FT intron 1475..1561
FT /*tag= c
FT intron 2174..2353
FT /*tag= d
FT intron 2488..2640
FT /*tag= e
XX
PN W08502610-A.
XX
PD 20-JUN-1985.
XX
PF 11-DEC-1984; 84WO-US02021.
XX
PR 30-NOV-1984; 84US-0675298.
PR 13-DEC-1983; 83US-0561024.
PR 21-FEB-1984; 84US-0582185.
PR 28-SEP-1984; 84US-0655841.
XX
PA (KIRI-) KIRIN-AMGEN INC.
XX

DR WPI; 1985-159229/26.
 DR P-PSDB; AAP50300.
 XX
 PT New polypeptide having properties of erythropoietin - is prepd.
 PT by cultivation of transformed eucaryotic or procaryotic host
 XX
 PS Disclosure; Page 43; 113pp; English.
 XX
 CC Human erythropoietin encoded by a sequence encoded by this phage
 CC lambda-hE1 is essential for red blood cell formation and is used
 CC for the diagnosis and treatment of blood disorders such as anaemia.
 CC Large amounts of EPO may be obtained using recombinant DNA
 CC techniques in contrast to small amounts obtained from plasma
 CC and urine. This sequence is expressed in E. coli. Bases indicated
 CC by a letter x were undetermined. See also AAN50345-6, AAN50348-50 and
 CC AAP50298-99, AAP50301.
 XX
 SQ Sequence 3211 BP; 658 A; 978 C; 929 G; 640 T; 6 other;

Query Match 38.5%; Score 15; DB 6; Length 3211;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 cctcaacccaggcgt 28
 |||||
 Db 248 cctcaacccaggcgt 262

RESULT 14
 AAV30956
 ID AAV30956 standard; DNA; 3211 BP.
 XX
 AC AAV30956;
 XX
 DT 11-SEP-1998 (first entry)
 XX
 DE Human erythropoietin encoding genomic DNA.
 XX
 KW Human; erythropoietin; EPO; bone marrow; reticulocyte; red blood cell;
 KW expression; CHO; chinese hamster ovary cell; diagnosis; blood disorder;
 KW ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 621..2643
 FT /*tag= a
 FT /product= "erythropoietin"
 FT /note= "contains introns"
 FT exon 621..633
 FT /*tag= b
 FT /number= 1
 FT intron 634..1072
 FT /*tag= c
 FT /number= 1
 FT exon 1073..1218


```

FT          /*tag= d
FT          /number= 2
FT  intron  1219..1474
FT          /*tag= e
FT          /number= 2
FT  exon    1475..1561
FT          /*tag= f
FT          /number= 3
FT  intron  1562..2173
FT          /*tag= g
FT          /number= 3
FT  exon    2174..2353
FT          /*tag= h
FT          /number= 4
FT  intron  2353..2487
FT          /*tag= i
FT          /number= 4
FT  exon    2488..6643
FT          /*tag= j
FT          /number= 5
XX
PN  AU688723-B.
XX
PD  19-FEB-1998.
XX
PF  02-DEC-1997;  97AU-0046867.
XX
PR  02-DEC-1997;  97AU-0046867.
XX
PA  (KIRI ) KIRIN AMGEN INC.
XX
PI  Lin F;
XX
DR  WPI; 1998-261957/24.
DR  P-PSDB; AAW58400.
XX
PT  Recombinant human erythropoietin - potentially useful for diagnosis
PT  and treatment of blood disorders
XX
PS  Example 5; Page 39-43; 100pp; English.
XX
CC  The present sequence encodes human erythropoietin (EPO), from an
CC  example of the present invention. The present invention describes
CC  recombinant human EPO which causes bone marrow cells to increase
CC  production of reticulocytes or red blood cells, where the polypeptide
CC  is the product of expression in CHO (Chinese hamster ovary) cells of
CC  an exogenous DNA sequence encoding human EPO. EPO is potentially
CC  useful in the diagnosis and treatment of blood disorders
CC  characterised by low or defective red blood cell production.
XX
SQ  Sequence 3211 BP; 658 A; 979 C; 928 G; 640 T; 6 other;

```

```

Query Match          38.5%;  Score 15;  DB 19;  Length 3211;
Best Local Similarity 100.0%;  Pred. No. 35;
Matches 15;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```

Qy 14 cctcaacccaggcgt 28
 |||||||
Db 248 cctcaacccaggcgt 262

RESULT 15

AAN60518

ID AAN60518 standard; DNA; 3401 BP.

XX

AC AAN60518;

XX

DT 01-JAN-1980 (first entry)

XX

DE Open reading frame coding for the erythropoietin tryptic fragment
DE of lambda HEP01.

XX

KW Erythropoietin; lambda HEP01; recombinant plasmid vector; anaemia;
KW mammal cell culture; 3T3; CHO; Chinese hamster ovary; ss.

XX

OS Homo sapiens.

XX

FH	Key	Location/Qualifiers
FT	intron	617..629
FT		/*tag= a
FT	intron	1195..1340
FT		/*tag= b
FT	intron	1598..1684
FT		/*tag= c
FT	intron	2296..2475
FT		/*tag= d
FT	intron	2610..2762
FT		/*tag= e

XX

PN WO8603520-A.

XX

PD 19-JUN-1986.

XX

PF 03-DEC-1985; 85WO-US02405.

XX

PR 22-JAN-1985; 85US-0693258.

PR 04-DEC-1984; 84US-0677813.

PR 03-JAN-1985; 85US-0688622.

XX

PA (GENE-) GENETICS INST INC.

PA (FRIT/) FRITSCH E.

XX

PI Fritsch E, Hewick RM, Jacobs K;

XX

DR WPI; 1986-169459/26.

DR P-PSDB; AAP60598.

XX

PT Prodn. of human cDNA clone expressing erythropoietin - for mass
PT prodn. of erythropoietin, useful for treating anaemia

XX

PS Disclosure; Page 19; 61pp; English.

XX

CC Recombinant plasmid vector lambda HEP01 expressing this genomic

CC fragment is expressed in e.g. 3T3 or CHO cell cultures. The
CC produced erythropoietin is useful for treatment of anaemia,
CC especially renal anaemia. The cloned gene expresses high levels of
CC the protein and thus provides a means of mass production. See
CC also AAN60513-17, AAN60519-21 and AAP60599.

XX

SQ Sequence 3401 BP; 698 A; 1033 C; 994 G; 676 T; 0 other;

Query Match 38.5%; Score 15; DB 7; Length 3401;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 cctcaacccaggcgt 28
|||||||
Db 240 cctcaacccaggcgt 254

Search completed: May 7, 2002, 19:18:17
Job time: 3810 sec

RESULT 5
 PAE18050
 LOCUS PAE18050 2310 bp DNA linear BCT 05-JUL-1999
 DEFINITION Pseudomonas aeruginosa intI1, blaVIM and aacA4 (partial) genes.
 ACCESSION Y18050
 VERSION Y18050.1 GI:5420397
 KEYWORDS aacA4 gene; aminoglucoside acetyl-transferase; beta-lactamase;
 blaVIM gene; DNA integrase; intI1 gene.
 SOURCE Pseudomonas aeruginosa.
 ORGANISM Pseudomonas aeruginosa
 Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 Pseudomonas.
 REFERENCE 1 (bases 1 to 2310)
 AUTHORS Lauretti, L., Riccio, M.L., Mazzariol, A., Cornaglia, G.,
 Amicosante, G., Fontana, R. and Rossolini, G.M.
 TITLE Cloning and characterization of blaVIM, a new integron-borne
 metallo-beta-lactamase gene from a Pseudomonas aeruginosa clinical
 isolate
 JOURNAL Antimicrob. Agents Chemother. 43 (7), 1584-1590 (1999)
 MEDLINE 99318582
 REFERENCE 2 (bases 1 to 2310)
 AUTHORS Rossolini, G.M.
 TITLE Direct Submission
 JOURNAL Submitted (04-SEP-1998) G.M. Rossolini, Dip. Biologia
 Molecolare-Sez., Microbiologia,, Univ. di Siena, via Laterina N.8,
 I- 53100 Siena, ITALY
 FEATURES
 source Location/Qualifiers
 1. .2310
 /organism="Pseudomonas aeruginosa"
 /strain="VR-143/97"
 /db_xref="taxon:287"
 /clone="pAC-2AL"
 gene complement(72. .1085)
 /gene="intI1"
 CDS complement(72. .1085)
 /gene="intI1"
 /codon_start=1
 /transl_table=11
 /product="DNA integrase"
 /protein_id="CAB46685.1"
 /db_xref="GI:5420398"
 /translation="MKTATAPLPPLRSVKVLDQLRERIRYLHYSLPTEQAYVHWVRAF
 IRFHGVRHPATLGSSEVEAFLSWLANERKVSVSTHRQALAAALFFYGKVLCTDLPWLQ
 EIGRPSPRRLLPVVLTPEVVRIILGFLEGEHRLFAQLLYGTCMRISEGLQLRVKDLDF
 DHGTIIIVREGKGSKDRALMLPESLAPSLREQLSRARAWWLKDAQEGRSGVALPDALER
 KYPRAGHSWPWFVFAQHTHSTDPRSGVVRHHMYDQTFQRAFKRAVEQAGITKPATP
 HTLRHSFATALLRSGYDIRTVQDLLGHSDVSTTMIYTHVLKVGAGVRSPLDALPPLT
 SER"
 gene 1122. .2141
 /gene="blaVIM"
 misc_feature 1122. .2135
 /gene="blaVIM"
 /note="gene cassette"
 misc_feature 1221. .1227
 /gene="blaVIM"
 /note="cassette upstream conserved recombination core"

maybe
 102 (a)
 size 51

```

CDS             site"
                1252. .2052
                /gene="blaVIM"
                /codon_start=1
                /transl_table=11
                /product="beta-lactamase VIM-1"
                /protein_id="CAB46686.1"
                /db_xref="GI:5420399"
                /translation="MLKVISSLLVYMTASVMAVASPLAHSGEPSGEYPTVNEIPVGEV
                RLYQIADGVWSHIATQSFDGAVYPSNGLIVRDGDELLIDTAWGAKNTAALLAEIEKQ
                IGLPVTRAVSTHFHDDRGGVDLRAAGVATYASPSTRRLAEAEAGNEIPTHSLEGLSS
                SGDAVRFGPVELFYPGAHSTDNLVVYVPSANVLYGGCAVHELSSSTSAGNVADADLAE
                WPTSVERIQKHYPEAEVVIPGHGLPGGLDLLQHTANVVKAHKNRSVAE"
misc_feature    2135. .2141
                /gene="blaVIM"
                /note="cassette downstream conserved recombination core
                site; cassette inverse recombination core site"
misc_feature    2136. .>2310
                /gene="aacA4"
                /note="gene cassette"
gene            2136. .2310
                /gene="aacA4"
CDS             2196. .>2310
                /gene="aacA4"
                /codon_start=1
                /transl_table=11
                /product="aminoglucoside acetyl-transferase"
                /protein_id="CAB46687.1"
                /db_xref="GI:5420400"
                /translation="MTEHDLAMLYEWLNRSHIVEWWGGEERPTLADVQEQY"
BASE COUNT     499 a      659 c      681 g      471 t
ORIGIN

```

```

Query Match          41.0%; Score 16; DB 1; Length 2310;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      14 cctcaactcaggcggtt 29
          |||||
Db      2122 CCTCAACTCAGGCGTT 2137

```

RESULT 9

HSPLGLN

LOCUS HSPLGLN 3490 bp mRNA linear PRI 14-DEC-1995

DEFINITION H.sapiens mRNA for plakoglobin.

ACCESSION Z68228

VERSION Z68228.1 GI:1122888

KEYWORDS plakoglobin.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 3490)

AUTHORS Franke,W.W., Goldschmidt,M.D., Zimbelmann,R., Mueller,H.M.,
Schiller,D.L. and Cowin,P.

TITLE Molecular cloning and amino acid sequence of human plakoglobin, the
common junctional plaque protein

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 86 (11), 4027-4031 (1989)

MEDLINE 89264555

REFERENCE 2 (bases 1 to 3490)

AUTHORS Zimbelmann,R.

TITLE Direct Submission

JOURNAL Submitted (14-DEC-1995) Zimbelmann R., German Cancer Research
Center, Division for Cell Biology, Im Neuenheimer Feld 280, D-69120
Heidelberg, Federal Republik of Germany

FEATURES Location/Qualifiers

source

1. .3490

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="HPG Ca 5.1"

/dev_stage="adult"

CDS

120. .2357

/codon_start=1

/product="plakoglobin"

/protein_id="CAA92522.1"

/db_xref="GI:1122889"

/db_xref="SPTREMBL:Q15151"

/translation="MEVMNLMEQPIKVTEWQQTYTYDSGIHSGANTCVPSVSSKGIME
EDEACGRQYTLKKTTTTYTQGVPPSQGDLEYQMSTTARAKRVREAMCPGVSGEDSSLLL
ATQVEGQATNLQRLAEPQLLKSIVHLINQDDAELATRALPELTKLLNDEDPVVVT
KAAMIVNQLSKKEASRRALMGSPQLVAAVVTRMQNTSDLDARCTTSILHNLSSHREG
LLAIFKSGGIPALVRMLSSPVESVLFYAITTLHNLLEYQEGAKMAVRLADGLQKMVPL
LNKNNPKFLAITTDCLQLLAYGNQESKLII LANGGPQALVQIMRNYSYEKLLWTTSRV
LKVLSVCPSNKP AIVEAGGMQALGKHLTSNSPRLVQNCLWTLRNLSDVATKQEGLESV
LKILVNQLSVDDVNVLTCATGTLNLTNNNSKNKTLVTQNSGVEALIHAILRAGDKDD
ITEPAVCALRHILTSRHPEAEMAQNSVRLNYGIPAIVKLLNQPNQWPLVKATIGLIRNL
ALCPANHAPLQEA AVIPRLVQLLVKAHQDAQRHVAAGTQQPYTDGVRMEEIVEGCTGA
LHILARDPMNRMEIFRLNTIPLFVQLLYSSVENIQRVAAGVLCELA QDKEAADAIDAE
GASAPLMELLHSRNEGATYAAAVLFRISEDKNPDYRKRVSVELTNSLFKHDPAAWEA
AQSMIPINEPYGDDMDATYRPMYSSDVPLDPLEMHMDMDGDYPIDTYS DGLRPPYPTA
DHMLA"

polyA_signal 3475. .3480

BASE COUNT 672 a 1172 c 979 g 667 t

ORIGIN

Query Match 41.0%; Score 16; DB 9; Length 3490;

Best Local Similarity 100.0%; Pred. No. 41;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 gctctgCGCCACCTCA 18
|||||||
Db 1491 GCTCTGCGCCACCTCA 1506

RESULT 10

G31640

LOCUS G31640 3490 bp DNA linear STS 28-SEP-1998

DEFINITION sWSS397 Eric D. Green Homo sapiens STS genomic, sequence tagged site.

ACCESSION G31640

VERSION G31640.1 GI:1916365

KEYWORDS STS.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 3490)

AUTHORS Bouffard,G.G., Iyer,L.M., Idol,J.R., Braden,V.V., Cunningham,A.F., Weintraub,L.A., Mohr-Tidwell,R.M., Peluso,D.C., Fulton,R.S., Leckie,M.P. and Green,E.D.

TITLE A collection of 1814 human chromosome 7-specific STSs

JOURNAL Genome Res. 7 (1), 59-64 (1997)

MEDLINE 97189344

PUBMED 9037602

REFERENCE 2 (bases 1 to 3490)

AUTHORS Green,E.D.

TITLE Human chromosome 7 STSs (1997)

JOURNAL Unpublished

COMMENT Synonyms: JUP

GDB: GDB:185732

GDB_DSEG: JUP

Contact: Eric D. Green

Genome Technology Branch

National Human Genome Research Institute/NIH

49 Convent Dr., MSC4431, Bldg. 49, Rm. 2A08, Bethesda, MD 20892

Tel: 3014020201

Fax: 3014024735

Email: egreen@nhgri.nih.gov

Primer A: GAGGCGTCGCGCGGGC

Primer B: GGTACAGGAGCAGGTTG

STS size: 253

PCR Profile:

Presoak: 0 degrees C for 0.00 minute(s)

Denaturation: 92 degrees C for 1.00 minute(s)

Annealing: 60 degrees C for 2.00 minute(s)

Polymerization: 72 degrees C for 2.00 minute(s)

PCR Cycles: 35

Thermal Cycler: PerkinElmer TC

Protocol:

Template: 30-100 ng

Primer: each 1 uM

dNTPs: each 200 uM

Taq Polymerase: 0.05 units/ul

Total Vol: 5 ul

Buffer:

MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3

This STS has been incorporated into the NHGRI chromosome 7 physical map, but was developed by another investigator. See GenBank record: Z68228 For additional information about the NHGRI chromosome 7 mapping project, see <http://www.nhgri.nih.gov/DIR/GTB/CHR7>. Also see Genomics 11:548-64 (1991) [MUID=92128937].

FEATURES
source Location/Qualifiers
1. .3490
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="7"
/clone_lib="Eric D. Green"
gene 1. .3490
/gene="JUP"
STS 636. .888
/gene="JUP"
primer_bind 636. .652
/gene="JUP"
primer_bind complement(872. .888)
BASE COUNT 672 a 1171 c 980 g 667 t
ORIGIN

Query Match 41.0%; Score 16; DB 11; Length 3490;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 gctctgCGCCACCTCA 18
|||||||
Db 1491 GCTCTGCGCCACCTCA 1506

(FILE 'HOME' ENTERED AT 14:51:31 ON 12 JUN 2002)

FILE 'MEDLINE, BIOSIS, CAPLUS' ENTERED AT 14:51:45 ON 12 JUN 2002
L1 66 S MUCIN8 OR MUC8 OR (MUCIN 8) OR (MUC 8)
L2 8 S L1 AND (GENOMIC OR CLON?)
L3 5 DUP REM L2 (3 DUPLICATES REMOVED)

(FILE 'HOME' ENTERED AT 18:16:14 ON 11 JUN 2002)

FILE 'MEDLINE, BIOSIS, CAPLUS' ENTERED AT 18:17:50 ON 11 JUN 2002

L1 32550 S MUCIN
L2 55 S L1 AND SPLICE
L3 28 DUP REM L2 (27 DUPLICATES REMOVED)

FILE 'STNGUIDE' ENTERED AT 18:19:44 ON 11 JUN 2002

FILE 'MEDLINE, BIOSIS, CAPLUS' ENTERED AT 18:29:16 ON 11 JUN 2002

L4 0 S L3 AND (MUC8 OR MUCIN 8)
L5 54 S (MUCIN 8) OR MUC8
L6 0 S L5 AND (VARIANT OR POLYMORPH? OR SNP)
L7 33 DUP REM L5 (21 DUPLICATES REMOVED)
L8 79 S RP11
L9 0 S RP11-0702?
L10 21 S L8(5A) (CLONE OR BAC)
L11 13 DUP REM L10 (8 DUPLICATES REMOVED)
L12 21 S ((BACTERIAL ARTIFICIAL) OR BAC) (3A) (CHROMOSOME(3A)7?)
L13 51 S ((BACTERIAL ARTIFICIAL) OR BAC) (3A) (CHROMOSOME(3A)7#####)
L14 16 DUP REM L13 (35 DUPLICATES REMOVED)
L15 24 S ((BACTERIAL ARTIFICIAL) OR BAC) (3A)
(CHROMOSOME(3A)12#####)
L16 14 DUP REM L15 (10 DUPLICATES REMOVED)